Synthesis of a *C*-Deoxyribonucleoside Modelled on Diaminopyrimidine: Evaluation of its Compatibility with the Modern DNA Alphabet

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Caroline Roost

von Thun / BE und Ermatingen / TG

Promotionskomitee:

Prof. Dr. Jay S. Siegel (Vorsitz)

Prof. Dr. Kim K. Baldridge

Prof. Dr. Roland K. O. Sigel

Zürich, 2012

Synthesis of a *C*-Deoxyribonucleoside Modelled on Diaminopyrimidine: Evaluation of its Compatibility with the Modern DNA Alphabet

ABSTRACT OF THE DISSERTATION

Synthesis of a *C*-Deoxyribonucleoside Modelled on Diaminopyrimidine: Evaluation of its Compatibility with the Modern DNA Alphabet

by

Caroline Roost

University of Zürich, 2012

Prof. Dr. Jay S. Siegel, Chair

From the beginning, mankind has had an interest in understanding who we are and from where we come. A chemical approach to answer these questions leads back in time to the point where the first living unit formed. Over the years several ideas of possible precursors of today's form of life have been discussed. Today's genetic storage material, deoxyribonucleic acid (DNA), consists of two different main groups of monomer units, which are pyrimidine- or purine-based nucleosides. According to a hypothesis that will be investigated in this thesis, an all-pyrimidine system existed on the early earth. All four nucleobases involved in the proposed pyrimidine system could have evolved from a single source, namely from 2,4-diaminopyrimidine. Simple hydrolysis could then have led to two modern pyrimidine bases, cytosine (C) and uracil (U), as well as to two exocyclic amino nucleosides (EAN's). These two EAN's are D and E, which are based on 2,4-diamiopyrimidine and cytosine, respectively. While D is presumably an analogue of the contemporary purine nucleoside adenosine (A), E can mimic guanosine (G).

In this work, model compounds for EANs were used due to stability reasons. By exchanging the exocyclic amino function by a carbon, the stability of these compounds could be increased. Homo-*C*-nucleoside D was successfully synthesised previously, although the proposed synthesis showed some drawbacks. The separation of four important intermediates was only achieved by high performance liquid chromatography (HPLC) separation, which is very costly, time-consuming and not suitable for the synthesis of larger amounts of material. In this thesis, several different syntheses and optimisations are investigated. The most promising synthetic pathway still requires a HPLC purification step. Nevertheless, the number of compounds that need to be isolated by HPLC was reduced from four to two. Ad-

ditionally, the total yield was increased. Furthermore, a successful way to obtain E, which could not have been synthesised beforehand, was proposed.

To investigate the base pairing properties of D, isothermal titration calorimetry (ITC) measurements were performed. The enthalpy of duplex formation from complementary oligonucleotide strands with varying middle positions were measured and compared. The results show a certain selectivity of D for U, suggesting that D is a mimic of A, as formulated by the hypothesis. Additionally, in one case, selectivity of D for A is observed. This can be explained by rotation of the nucleobase around the methylene bridge of the homo-*C*-nucleoside.

ZUSAMMENFASSUNG

Synthesis of a *C*-Deoxyribonucleoside Modelled on Diaminopyrimidine: Evaluation of its Compatibility with the Modern DNA Alphabet

von

Caroline Roost

Universität Zürich, 2012

Prof. Dr. Jay S. Siegel, Vorsitz

Seit jeher hat sich die Menschheit dafür interessiert, wer wir sind und woher wir kommen. Chemiker versuchen dieser Frage nachzugehen indem sie den Zeitpunkt, an dem die erste lebende Einheit entstand, erforschen. Über die Jahre wurden verschiedene mögliche Vorläufer unserer heutigen Lebensform diskutiert. Die heutige Erbsubstanz, Deoxyribonukleinsäuren (DNS), besteht aus zwei verschiedenen Arten von Einzeleinheiten. Dies sind Pyrimidine- oder Purin-basierte Nukleoside. Einer Hypothese folgend, welche in dieser Doktorarbeit eingehend untersucht wird, existierte einst ein genetischer Code, welcher nur aus Pyrimidine-Nukleosiden bestand. Alle vier Nukleobasen, welche in diesem vorgeschlagenen Pyrimidinsystem existierten, könnten von einer gemeinsamen Quelle abstammen, nämlich 2,4-Diaminopyrimidin. Durch Hydrolyse bildeten sich dann die heute vorkommenden Pyrimidinbasen Cytosin (C) und Uracil (U) und zudem zwei exozyklische Aminonukleoside (EANs). Diese beiden EANs sind D und E, welche auf 2,4-Diaminopyrimidin bzw. Cytosin basieren. D ist ein Analog von Adenosin (A), während E ein Analog von Guanosin (G) ist.

Aus Stabilitätsgründen wurde in dieser Arbeit mit Modelverbindungen für die EANs gearbeitet. Durch Austausch des exozyklischen Stickstoffs mit Kohlenstoff konnte die Stabilität dieser Verbindungen erhöht werden. Homo-*C* Nukleosid D konnte vorgängig bereits erfolgreich hergestellt werden, aber die vorgeschlagene Synthese hatte einige Nachteile. Die Trennung von vier wichtigen Zwischenstufen konnte nur mittels Hochleistungs-Flüssigchromatographie (HPLC) erreicht werden. Diese Methode ist sehr teuer, zeitintensiv und ungeeignet für grössere Mengen von Material. In der vorliegenden Arbeit wurden mehrere verschiedene Synthesen und Optimierungen untersucht. Der aussichtsreichste Pfad

braucht immer noch einen HPLC-Reinigungsschritt, aber die Anzahl Verbindungen, welche per HPLC getrennt werden müssen, konnte von vier auf zwei reduziert werden. Zudem konnte die Gesamtausbeute erhöht werden. Ausserdem wurde ein Weg zur Synthese von E gefunden, welches vorgängig nicht hergestellt werden konnte.

Zur Untersuchung der Eigenschaften zur Basenpaarung von D wurden isothermale Titrationskalorimetrie-Messungen (ITC) durchgeführt. Die Enthalpie der Duplexausbildung von komplementären Oligonukleotidsträngen mit variabler Mittelposition wurde gemessen und die Resultate miteinander verglichen. Die Auswertung zeigt eine gewisse Selektivität von D für U. Dies lässt darauf schliessen, dass D ein Analog von A ist, wie es die untersuchte Hypothese besagt. Zudem kann in einem Fall eine Selektivität von D für A beobachtet werden. Dies kann mit einer Drehung der Nukleobase um die Methylbrücke des Homo-*C* Nukleosids erklärt werden.

Table of Contents

List of Figures

List of Schemes

List of Tables

1 Introduction

1.1 Origin of Life

Ever since the beginning of time, humans have had an interest in understanding where we come from and, consecutively, who we are. The first attempts to address those questions were of philosophical nature. Over the past few decades, chemists started to take an interest in the questions concerning life and began postulating how the very first life on earth formed, referring to this time as the origin of life. To address these questions about life, a definition of the term "life" is needed but opinions differ highly on this question. The most rudimentary definition of a "living unit" requires the capability of reproduction, therefore including a system that stores and transmits genetic information. In all present-day organisms, except for some viruses, this storage system of genetic information is deoxyribonucleic acid (DNA).[1]

There are two main questions in the field of origin of life chemistry. The first question asks which biopolymer was first, DNA, ribonucleic acid (RNA) or peptides, and from which precursor it evolved. The second question centres around the problem of compartmentalisation, meaning the organisation of a living unit in different compartments and the transmission of genetic material from one compartment to an other. In this project, manly the first question about the precursor of the first biopolymer is addressed.

One big difficulty in finding answers regarding the origin of life is the fact that we cannot be sure about the conditions found on the prebiotic earth. Generally, a moderate pH and aqueous environment is assumed. *Miller* and *Urey* proposed that the earth started with a reducing atmosphere that became later oxidising due to the escape of hydrogen.[2] The largest source of uncertainty in this prebiotic atmosphere is the temperature. A hydrothermal environment with up to 100 °C has been proposed due to the fact that thermophilic microorganisms, which grow at elevated temperatures, are found to be the oldest known organisms to date.[3] Other experts like *Bada* and *Lazcano* have an opposite theory. They assume that a low temperature environment existed due to the instability of biomolecules at high temperatures and the fact that mineral-based concentration of molecules, involving weak noncovalent bonds, is most effective at low temperatures.[4]

Regardless of the exact conditions at this time on Earth, the primordial soup theory is nonetheless widely accepted. Independently developed by *Oparin* and *Haldane*[5] in the 1920s,

a modern version of the theory states that organic compounds accumulated in the primordial oceans and underwent polymerisation, thereby produced complex macromolecules that eventually developed the ability to catalyse their own replication, thus leading to the origin of life, the RNA world and finally to today's DNA/RNA/Protein system (*Figure 1.1*).

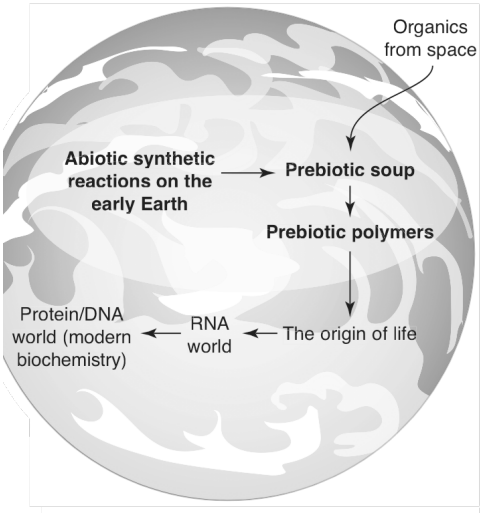


***Figure 1.1**: Various steps from the formation of the primordial soup to the origin of life on the early earth.*[4]

1.1.1 Historical Overview

Present models state that the formation of the earth began approximately 4.5 billion years ago and that the first living organisms, prokaryotic microfossils, are around 3.6 billion years old.[6,7] Between these two mile stones lies the time interval of prebiotic chemistry, followed by the RNA world, that will be described in *Section 1.1.2*.

An important impulsion in the field of prebiotic chemistry was the famous experiment by *Miller* and *Urey* in 1953.[2,8] Based on the primordial soup theory, the researchers mimicked the atmosphere on the early Earth by exposing CH_4, NH_3, H_2O and H_2 to an electric discharge in a closed circular system (*Figure 1.2*) and they were able to isolate amino acids in the milligram range. For a long time after this experiment, it was believed that proteins were the first biopolymers on earth and therefore the prebiotic hereditary material. This assumption was based on the results of the *Miller-Urey* experiment, where amino acids were synthesised under presumably prebiotic conditions, as well as based on the fact that proteins have been

known to possess catalytic activities. Often, a co-existence of proteins and RNA was assumed, with the latter holding the genetic information.[9]

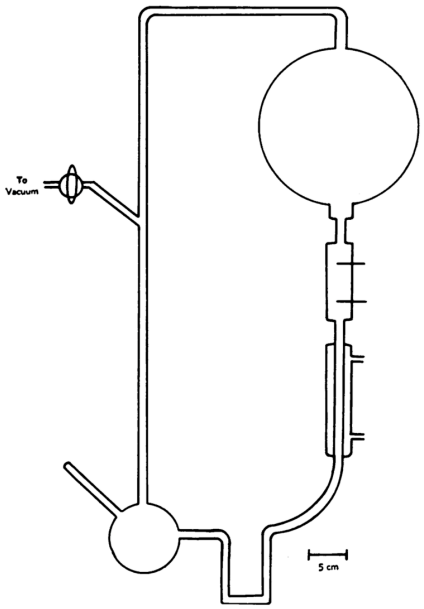


***Figure 1.2**: Experimental setup of the Urey-Miller experiment.*[8]

It was not until the 1980s, when both, *Cech* and co-workers and *Altman* and co-workers independently found that RNA exhibited catalytic activities. *Altman* and his team worked with *Escherichia coli* and found that ribonuclease-P cleaves phosphodiester bonds during the maturation of the transfer RNA (tRNA) molecule.[10,11] *Cech et al.* discovered that the ribosomal RNA (rRNA) of *Tetrahymena* contained a self-splicing exon.[12,13] In 1989, *Cech* and *Altman* were jointly awarded the *Nobel Prize in Chemistry* for their discovery of catalytic properties of RNA. From this discovery, RNA appeared to be the long-sought piece that combines the ability to store genetic information while presumably catalysing its own replication. Prior to this discovery it had been speculated that RNA is both the original genetic as well as catalytic material[14,15] but only the finding of the catalytic properties of RNA opened the way for the RNA world hypothesis, which will be described in detail in the following section.

1.1.2 RNA World

The RNA-world hypothesis is currently the most accepted theory of the origin of biopolymers and life, even it is solely based on indications. The hypothesis states that at the

beginning of chemical evolution there were only RNA molecules, since RNA provided all of the functions needed for the first biopolymers to evolve: providing genetic information and serving to catalyse the synthesis of themselves.[16] Furthermore, it is the connection between DNA and proteins, thus between genetic material and metabolism. According to the RNA-world hypothesis, at the first stage of chemical evolution, RNA performed the catalytic activities to assemble themselves from the primordial soup. By using recombination and mutation, they could developed new functions and adapt to new niches. Using RNA cofactors then have led to an entire range of enzymatic activities.[14] Later, RNA started to synthesise proteins, first by developing RNA adapters that bound amino acids and then by arranging them according to an RNA template using other RNA molecules. This process provided the first proteins, which were able to carry out the same enzymatic reactions like RNA but more rapidly and more effectively. At the next stage DNA appeared, presumably copied from the RNA molecules by reverse transcription. With double-stranded DNA, a stable genetic information store was formed that possessed error-correcting properties because of its double-stranded structure but was still capable of mutation and recombination.[9] Following the widely accepted RNA-world hypothesis, that is how the present genetic apparatus has evolved.

To accept that RNA might have been the primordial hereditary material, one should be able to show its synthesis under prebiotic conditions. This challenge was not fulfilled until *Oró* published a purine synthesis under prebiotic conditions. As a key intermediate, *Oró* used hydrogen cyanide, a reagent that was also isolated in the *Miller-Urey* experiment.[17] The prebiotic synthesis of pyrimidine turned out to be more difficult than purine. The first synthesis of cytosine was proposed by *Orgel* and his team. They were able to synthesise cytosine under prebiotic conditions from cyanoethylene and cyanate but only in very poor yields.[18] Despite several optimisation attempts the yields could not be remarkably improved.[19] Several decades later, *Robertson* and *Miller* finally managed to synthesise cytosine under prebiotic conditions in up to 50 % yield. They used a pathway starting from cyanoacetaldehyde (formed by hydrolysis of cyanoacetylene) and let it react with a concentrated urea solution. Uracil is easily formed by hydrolysis of cytosine.[20]

Even after the synthesis of purine and pyrimidine, an important question was still unanswered: How did the ribonucleotides form? The most obvious approach would be a condensation of the furanose form of ribose and the nucleobase. The problem with this approach is that it only works inefficiently in the case of purines[21] and simply does not work for pyrimidines[22] due to a kinetic reason (the N1 lone pair of cytosine and uracil is unavailable because of delocalisation) and, in water, a thermodynamic reason (the equilibrium constant is such that hydrolysis of ribonucleoside to pyrimidine and ribose is favoured over condensa-

tion). It was not until recently, that *Sutherland* and co-workers proposed a way to synthesise pyrimidine ribonucleotides under prebiotic conditions.[23] Their elegant synthesis starts from cyanamide, cyanoacetylene, glycoaldehyde, glyceraldehyde and inorganic phosphate, which are all plausible prebiotic molecules.[18,24] The synthetic pathway bypasses free ribose and pyrimidines and proceeds instead through arabinose amino-oxazoline (*Figure 1.3*). The synthesis of *Sutherland* and his team was highly appreciated in the origin of life society and gave another boost to the RNA world hypothesis.

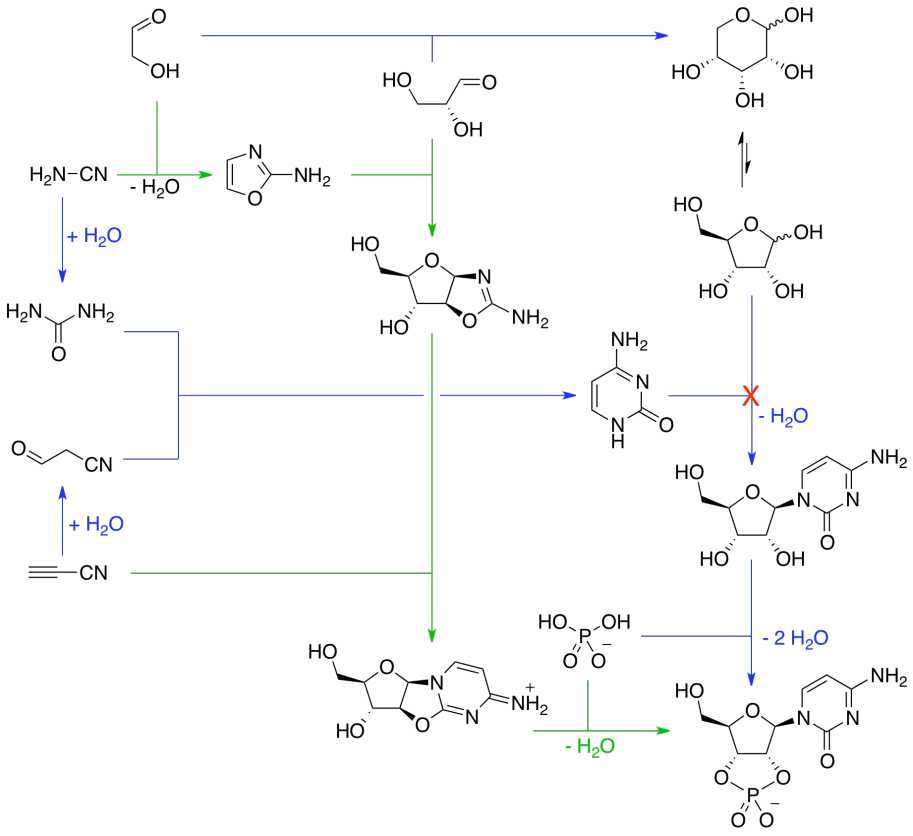


***Figure 1.3**: Pyrimidine ribonucleotide assembly options.*[23]

If there was a RNA world, there must also have been a time period before, known as the pre-RNA world. It is assumed that in this pre-RNA world, living organisms contained a backbone different from ribose-phosphate and also bases different from C, G, A and U. RNA could then have evolved from these precursors.[25] In the next section, such possible precursors will be presented.

1.1.3 Precursors of present-day DNA/RNA

Two requirements for a possible precursor of DNA and RNA are that it should be stable and synthetically accessible under prebiotic conditions. To find such a progenitor, there are three different parts to alter: the sugar moiety, the phosphate backbone and the nucleobase.

The group of *Eschenmoser* followed the question, 'why did nature chose ribofuranosyl nucleic acids instead of another sugar family' and studied therefore several sugar modifications under a prebiotic point of view. The most famous example of such a modification is the homo-DNA, bearing a six-membered pyranose ring instead of the present five-membered ribose (*Figure 1.4*, left).[26,27] Presumably due to its higher rigidity, homo-DNA forms stronger base pairs than natural DNA but does not cross-pair with natural nucleic acids. Therefore, homo-DNA was shown to be especially suitable as molecular beacons.[28] Another interesting example of sugar modification by *Eschenmoser et al.* is the α-L-threofuranosyl nucleic acid (TNA).[29] It contains tetrafuranose instead of D-ribose and is linked in a 3'→2' manner (*Figure 1.4*, middle). Surprisingly, TNA shows strong cross-pairing abilities towards its natural counterparts despite their conformational and constitutional differences.

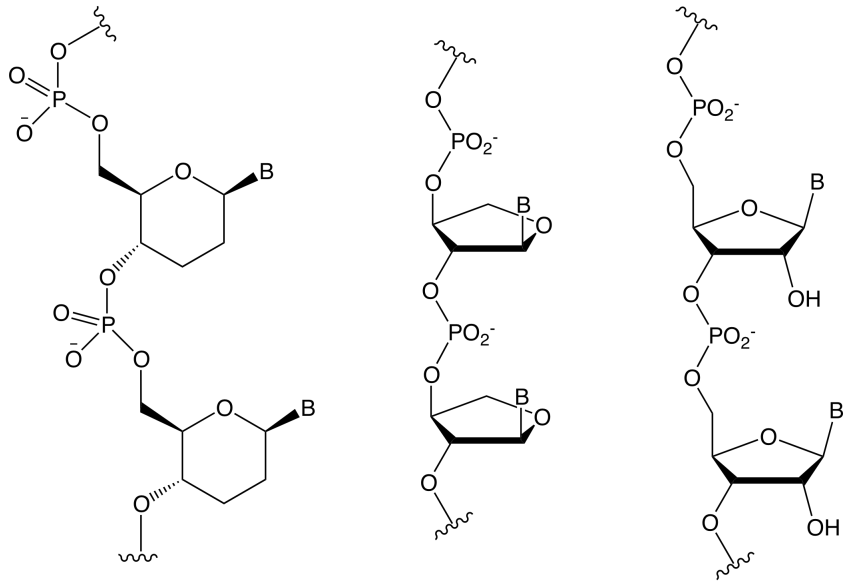


***Figure 1.4**: Structure of Homo-DNA (left) and TNA (middle) compared to RNA (right).*

For backbone modifications, a variety of examples exists. *Ueda* and co-workers were the first to propose an acyclic backbone called glycerol nucleic acid (GNA, shown in *Figure*

1.5, left).[30] Later, *Orgel et al.* found that GNA exhibits stronger *Watson-Crick* pairing than natural DNA or RNA, despite the fact that GNA consists of acyclic three carbon backbone units.[31] *Miller* and co-workers proposed another backbone modification. They suggested a peptide nucleic acid (PNA), similar to proteins but with bases contained in RNA instead of amino acids in the side chains (*Figure 1.5*, middle). PNA can be synthesised under prebiotic conditions.[32] The idea of *Miller et al.* is based on work by *Westheimer*, who pursued the question 'why does nature chose phosphates for the backbone in nucleic acids?' and came to the conclusion that phosphates meet all of the required conditions and that no alternative is obvious.[33] *Usher* worked on another interesting question: why does RNA have a 3',5' linkage? In theory, a linkage through the 2'-OH would be possible but *Usher* concludes that the 3',5' linkage serves as a protection against internal strand breakage.[34]

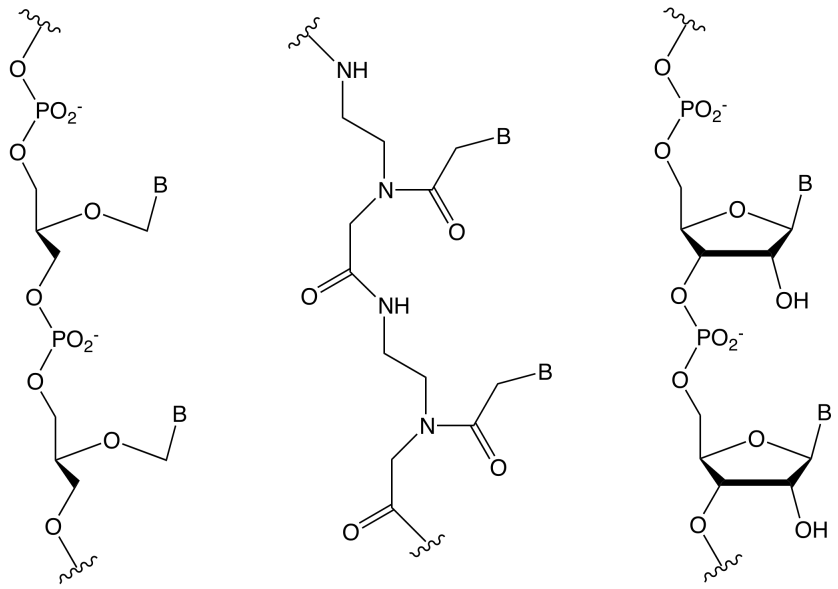


Figure 1.5*: GNA (left) and PNA (middle) compared to RNA (right).*

Some examples for the third site of evolutionary modification, the nucleobase, are found in literature. *Benner et al.* expanded the genetic alphabet by proposing several new base pairs (*Figure 1.6,* top).[35] This expansion could have led to a greater diversity in structure and functionality and therefore greater susceptibility to evolution. The described building blocks were not synthesised under prebiotic conditions, though. In contrast, *Miller* and co-workers proposed an urazole based precursor that could be synthesised under early-Earth conditions.[36] Urazole forms hydrogen bonds with adenine (*Figure 1.6*, bottom), comparable with those of

uracil. While uracil is completely unreactive with ribose under mild aqueous conditions, urazole reacts spontaneously with ribose and could therefore have been a precursor of uracil.

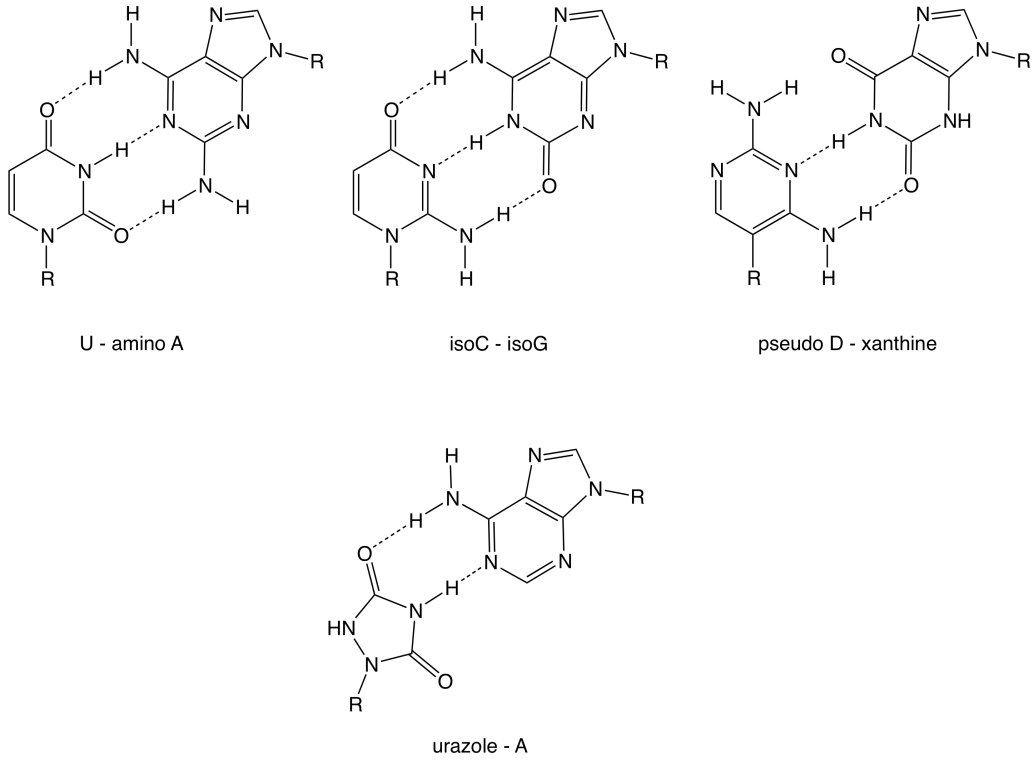


***Figure 1.6**: Alternative base pairs.*

In 1988, when the synthesis of pyrimidine nucleosides under prebiotic conditions was still unrevealed, *Wächtershäuser* proposed an all-purine base pairing code.[37] His assumption was based on the fact that pyrimidines cannot oligomerise onto a complementary template without assistance of a polymerase due to their low stacking energies. However, purines with their higher stacking energy do oligomerise.[38]

A counter hypothesis was proposed by *Siegel* and *Tor*.[39] They suggested an all-pyrimidine code with three bases, including 2,4-diamionopyrimidine and the naturally occurring cytosine and uracil. 2,4-diaminopyrimidine can be synthesised under prebiotic conditions and by simple hydrolysis, it can form cytosine and uracil. Thus, it fulfils the single source criterion. This hypothesis is described in depth in *Section 2.1*.

1.2 DNA Structure

1.2.1 Historical Overview

In 1866, *Haeckel* proposed that the material responsible for hereditary properties is located in the nucleus of a cell.[40] At that time proteins were considered to contain all hereditary traits since it was the only cell material known that showed a sufficient complexity to fulfil this task. Three years later, *Miescher* succeeded in the first isolation of DNA from leukocytes of pus, found on discarded surgical bandages.[41,42] Throughout his experiments, he noticed that a substance precipitated from the solution upon adding acid and dissolved again upon adding alkali. What he described was the first crude precipitation of DNA. *Miescher* named the isolated material *nuclein*, since he found it in the nuclei of the cells. Later tests by *Miescher* showed that *nuclein* lacked sulfur but contained a large amount of phosphorus.[42] This showed the new material to be significantly different from proteins. *Boveri* postulated in 1902, that the hereditary units were located on chromosomes[43] and seven years later, *Johannsen* termed these hereditary units *genes*.[44] In 1928, *Griffith* postulated the *transforming principle*.[45] He discovered that injection of a mixture of heat-killed virulent S pneumococci (*Diplococcus pneumoniae*) and non-pathogenic R pneumococci in mice resulted in death of a majority of the mice. Since S pneumococci were found in the blood of the dead mice, Griffith concluded that a transformation of the R to the S form occurred, induced by the heat-killed S pneumococci. *Levene* identified the four building blocks of DNA, including the four bases adenine, cytosine, guanine and thymine.[46,47] Since he found all four bases in an equal ratio, he proposed the tetranucleotide hypothesis, postulating a monotonously repeating sequence (*Figure 1.7*), which was therefore not suspected to have any genetic function.

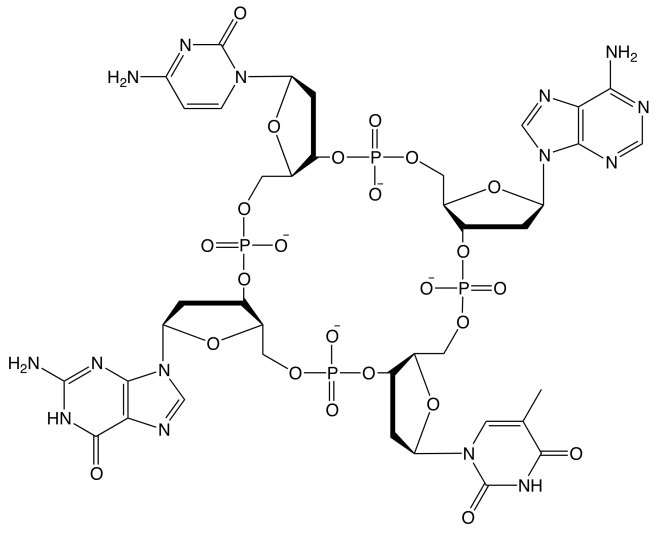


Figure 1.7: *Hypothetical tetranucleotide structure.*

In 1944, *Avery*, *MacLeod* and *McCarty* demonstrated that *Griffith's transforming principle* is DNA.[48] They showed this transforming principle to be unaffected by enzymes that catalyse the hydrolysis of proteins and RNA, such as trypsin and ribonuclease, but to be completely inactivated by DNase. Additionally, the transforming principle showed all the physical and chemical properties of DNA but contained no detectable protein. At the end of the 1940s, *Chargaff* determined that within a species, the bases in DNA were always present in fixed ratios and he could thereby refute the tetranucleotide hypothesis and at the same time, indicated that DNA could have a sufficient complexity for fulfilling the requirements of a genetic material.[49]

The structure of DNA was finally revealed in the 1950s. *Franklin* and *Wilkins* demonstrated the regularly repeating helical structure of DNA based on X-ray analyses.[50,51] With help of this X-ray analyses, *Watson* and *Crick* proposed in 1953 the molecular structure of DNA to be a double helix in which adenosine always pairs with thymidine and cytidine always pairs with guanosine.[52] An additional important milestone in the history of DNA research was the development of *Polymerase Chain Reaction* (PCR) by *Mullis* in 1983, which eased the synthesis of DNA enormously and allowed the production of large amounts of DNA in a short time.[53] Finally, in 1990, the sequencing of the human genome begun and it took over a decade, until 2001, when the complete sequence could finally be published.[54] Today, a lot of different functions of DNA are known. It not only holds the central role in storage and expression of genetic information but nucleosides and nucleotides also participate in many biochemical processes, like energy-releasing pathways and enzymatic reactions.[1]

1.2.2 Nucleosides and Nucleotides

Nucleosides are the monomer units of nucleic acid. Nucleotides are the phosphate esters of nucleosides. They consist of three main parts: a sugar moiety, a nucleobase and a phosphate group (*Figure 1.8*).

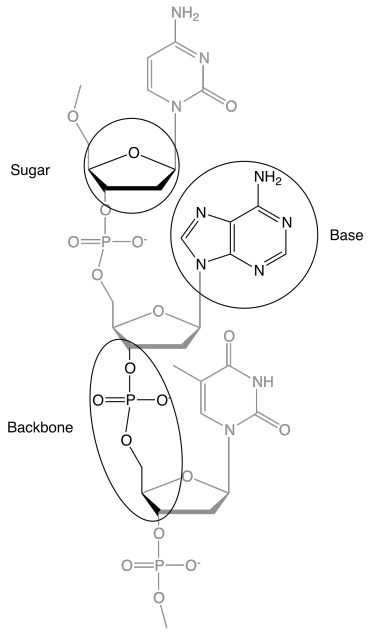


***Figure 1.8**: Three main components of nucleic acids: base, sugar and phosphate.*

The nucleobase is a nitrogen-containing heterocyclic base, either a purine in the case of adenine and guanine or a pyrimidine in the case of cytosine, uracil and thymine. The two purine bases as well as the pyrimidine base cytosine are found as nucleobases in both RNA and DNA. The fourth base occurring in RNA is uracil, while in DNA it is thymine (5-methyluracil). *Figure 1.9* shows the structure of the five bases.

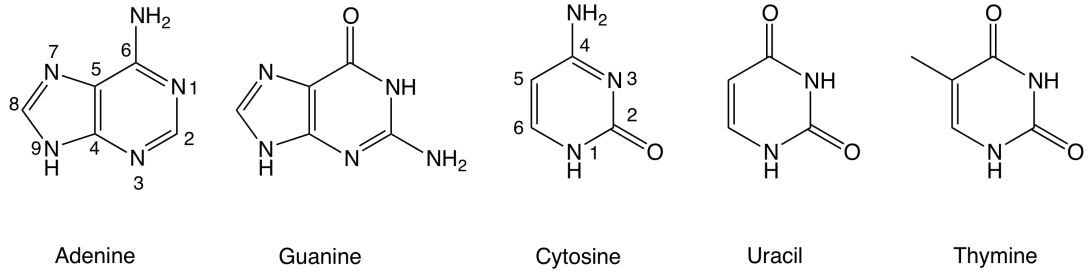


***Figure 1.9**: Chemical structure and numbering of the nucleobases.*

The nucleosides are named after the corresponding nucleobase that they contain (*Figure 1.10*). In the case of adenosine (A) and guanosine (G), the N9 is linked to the C1' of the pentose, while for cytidine (C), uridine (U) and thymidine (T), it is the N1 that is linked to the C1' of the sugar. This linkage is called the glycosidic bond. In naturally occurring RNA and DNA, it is in its β-configuration, meaning that the glycosidic bond is on the same side of the sugar like the C4'-C5' bond. If these two bonds are on opposite sides of the sugar, the configuration is a α-configuration.

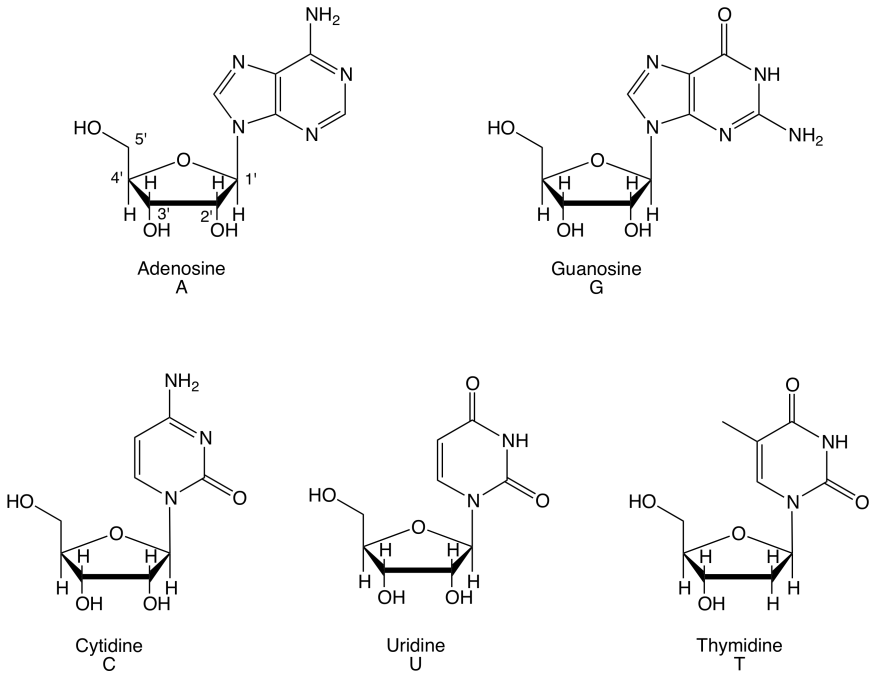


***Figure 1.10**: Chemical structure and numbering of nucleosides.*

The sugar moiety is a pentose sugar, D-ribose in the case of RNA and 2-deoxy-D-ribose in the case of DNA. With deoxyribose, the name of the nucleoside changes for example in the case of adenosine to deoxyadenosine (see *Table 1.1*). The pentose sugar is locked in a five-membered ring.

The phosphate group can be bonded to the C5'-hydroxygroup (5'-nucleotide) or to the C3'-hydroxygroup (3'-nucleotide). The name changes then for example from adenosine to adenylic acid (see *Table 1.1*).[55]

Symbol	Base	Nucleoside	Nucleotide
		Ribonucleosides and -nucleotides	
A	Adenine	Adenosine	Adenylic acid
G	Guanine	Guanosine	Guanylic acid
C	Cytosine	Cytidine	Cytidylic acid
U	Uracil	Uridine	Uridylic acid
		Deoxyribonucleosides and -nucleotides	
A	Adenine	Deoxyadenosine	Deoxyadenylic acid
G	Guanine	Deoxyguanosine	Deoxyguanylic acid
C	Cytosine	Deoxycytidine	Deoxycytidylic acid
T	Thymine	Thymidine	Thymidylic acid

***Table 1.1**: Nomenclature of nucleosides and nucleotides.*

1.2.3 Physical Properties

1.2.3.1 Tautomerism

In solution, heterocyclic molecules usually exist in a rapid equilibrium of tautomers due to hydrogen atoms attached to nitrogen atoms, which are able to migrate to other free nitrogens or to keto oxygens within the same molecule.[56] The tautomerism mainly depends on the electric constant of the solvent and on the pK_a of the heteroatoms. In general, the naturally occurring bases are predominantly in the keto and amino tautomeric forms with an excess of 10^3-10^6 over the enol and imine forms, respectively.[57] This is essential, since uracil and guanine in the enol form simulate cytosine and adenine and cytosine and adenine in the imino form may substitute for uracil and guanine (*Figure 1.11*). This substitution may lead to point mutations if not detected by repair enzymes.[55]

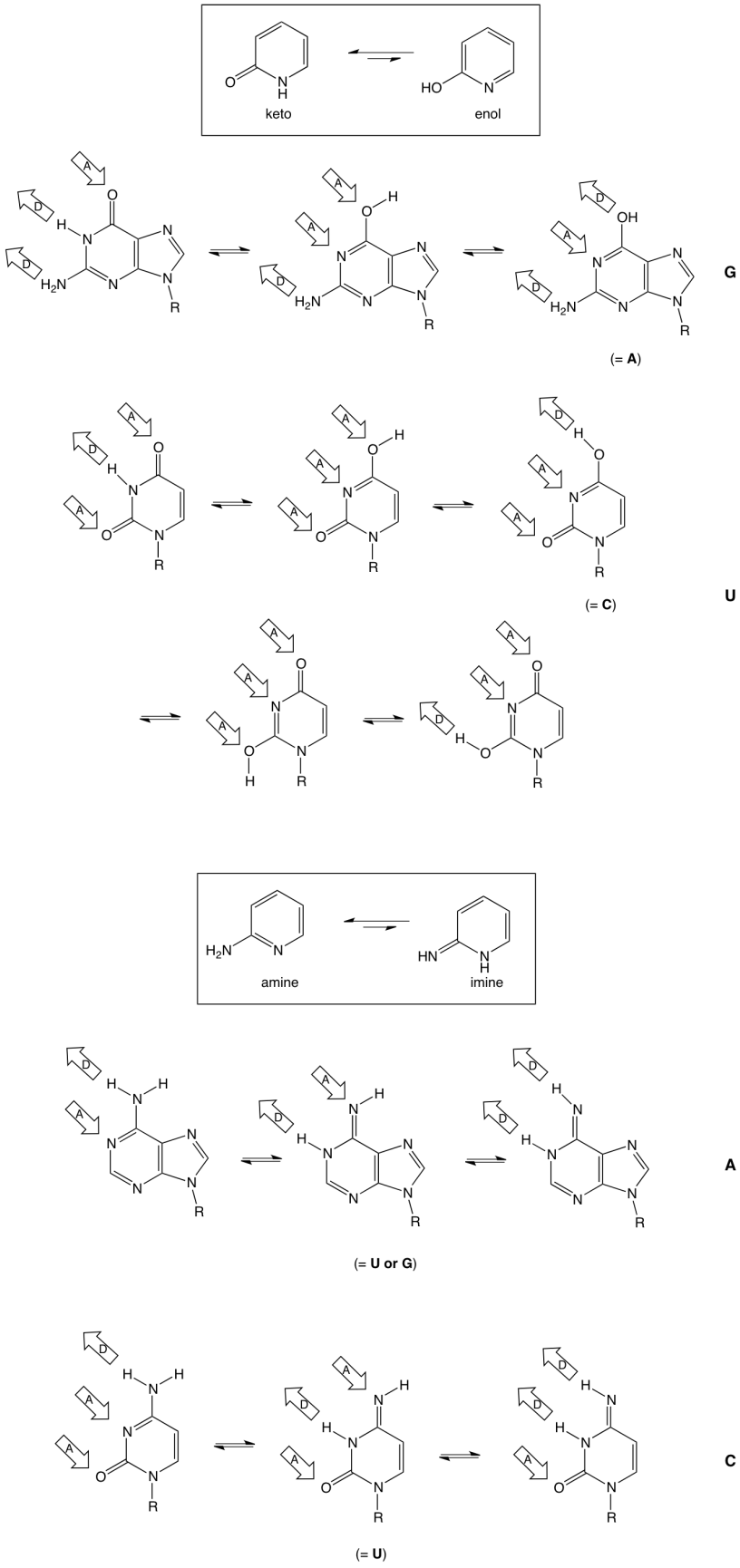


***Figure 1.11**: Keto-enol tautomerism of G and U and amine-imine tautomerism of A and C.*

1.2.3.2 Protonation (pK_a)

The acid-base behaviour is an important physical characteristic of nucleosides since protonation of the nucleobase changes the acceptor/donor abilities and therefore the base pairing properties. Protonation and deprotonation of the base occurs between pH 3 and 10. pK_a values for bases in ribonucleosides and 2'-deoxyribonucleosides are shown in *Table 1.2*.

Compound	pK_a	
(Site of protonation)	Deoxynucleoside	Nucleoside
Adenin (N1)	3.5	3.5
Guanin (N7)	2.3	2.1
(N1)	9.5	9.6
Cytosine (N3)	4.2	4.0
Thymine (N3)	9.7	9.9
Uracil (N3)	9.5	9.6

***Table 1.2**: pK_a values of bases in 2'-deoxyribonucleosides and ribonucleosides (298 K).*[58]

The site of protonation is in all cases the ring nitrogen, not the exocyclic amino nitrogen. Although the amino nitrogens have greater electron density, the ring nitrogens are more acidic due to better stabilisation of the protonated species by resonance.[55] The pK_a values for the 2'-hydroxy groups of ribonucleosides A, G, C and U were found to be 12.2, 12.5, 12.5 and 12.6, respectively.[59]

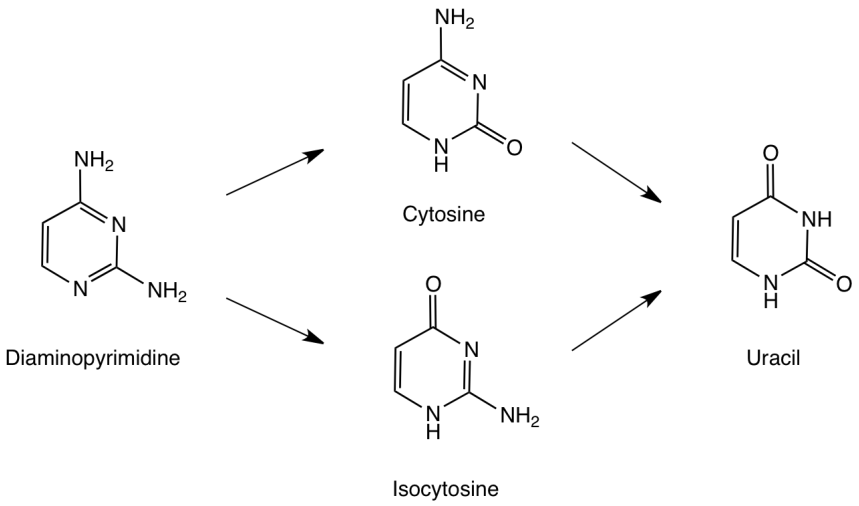
1.2.3.3 Stability

Stability is a crucial point in origin of life science since for a compound to be used in the first living organism it needs to show sufficient stability in order to keep the balance between synthesis and degradation. A high degradation rate would lead to vanishingly small concentrations. Especially the stability at high temperatures is of interest since a high-temperature-origin of life with temperatures between 80 and 110 °C is widely favoured (*see Section 1.1*). A common source of instability in nucleosides is the nucleobase, especially in form of deamination of the base.[60] By deamination, guanine can be converted to xanthine, adenine to hypoxanthine and cytosine to uracil. Especially the latter is of importance for this project. *Table 1.3* shows the half-lives of the naturally occurring nucleobases and diaminopyrimidine (D) in aq. solution ($2.5 \cdot 10^{-5}$ to $1 \cdot 10^{-3}$ M concentration in 0.05 M phosphate buffer; pH 7).[61]

Temperature	A	G	C	T	U	D
0 °C	$6.0 \cdot 10^5$ yr	$1.3 \cdot 10^6$ yr	17 000 yr	$2.0 \cdot 10^9$ yr	$3.8 \cdot 10^8$ yr	40 000 yr
25 °C	10 000 yr	10 000 yr	340 yr			
100 °C	1 yr	0.8 yr	19 d	56 yr	12 yr	42 d

***Table 1.3**: Half-lives of the naturally occurring nucloebases and of diaminopyrimidine.*

The most stable bases are thymine and uracil, followed by the purine bases adenine and guanine. By far the shortest half-life has cytosine. Even the half-life of 17,000 years at 0 °C is very short on a geologic time scale. Because 2,4-diaminopyrimidine is a centrepiece of this project, its stability is also discussed here. The half-live of D is comparable to the one of C. It can undergo hydrolysis to cytosine and isocytosine and further to uracil (*Scheme 1.1*).



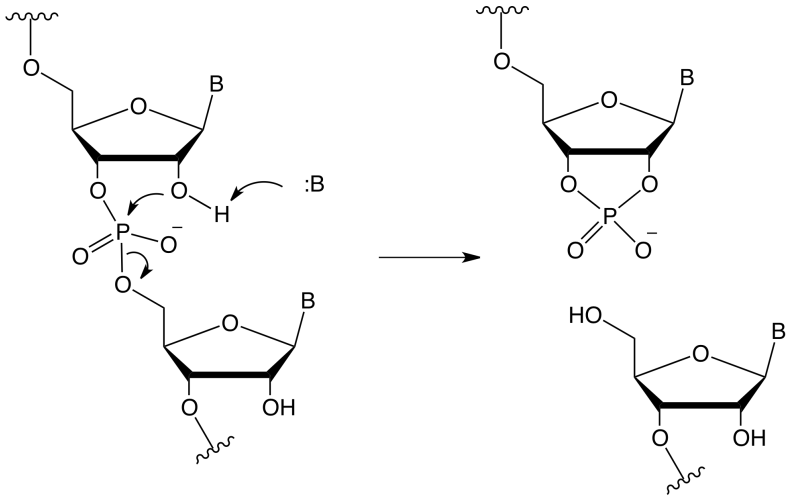
***Scheme 1.1**: Hydrolysis of diaminopyrimidine.*

The stability for nucleosides and nucleotides in single stand DNA does not remarkably change. The half-life for cytidine in single-stranded DNA is 200 years at 37 °C, which is about the same as the free cytosine base. However, in double-stranded DNA the half-life increases to 30,000 years, which is a 150-fold increase. This leads to the conclusion that deamination of the nucloebase is strongly dependent on the environment.[62]

Nucleobases are not the only parts of nucleotides that are susceptible to decomposition or cleavage. Another vulnerable point of attack is the *N*-glycosidic bond. Hydrolysis of this bond leads to a release of free base. In general, purine nucleotides release their base 20 times

faster than pyrimidine nucleotides. Interestingly, the rate of depurination of double-stranded DNA is only slowed down four times compared to single-stranded DNA.[63]

Another point that should be mentioned here is the difference in stability between RNA and DNA. The 2'-OH group of RNA stabilises the *N*-glycosidic bond. However, it makes the phosphodiester bond much more susceptible to hydrolysis (*Scheme 1.2*), especially in the presence of base.[64]



***Scheme 1.2**: Hydrolysis of RNA.*

1.2.4 Nucleic Acids

Nucleotide chains containing 20 or less monomer units are called oligonucleotides. Nucleic acids are defined as chains with more than 20 nucleoside units. There are two main classes of nucleic acids: ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). They differ in the sugar moiety, which is a D-ribose in case of RNA and a 2-deoxy-D-ribose in case of DNA.

1.2.4.1 Primary Structure

In natural DNA, the nucleotides are always linked by a phosphodiester bond between the 5'-end of one nucleotide to the 3'-end of another. There are no 5'-5' or 3'-3' linkages, which means that the uniqueness of a DNA primary structure only depends on the sequence of its bases.[65]

The sugar-phosphate structure of a nucleotide in a nucleic acid is defined by the torsion angles α, β, γ, δ, ε and ζ while the glycosidic bond is described by χ (*Figure 1.12*, left). Despite these numerous torsion angles, nucleic acids are not as flexible as one would imagine. The torsion angles are subjected to many intramolecular restraints. For example about the glycosidic bond, only two orientations of the base are sterically reasonable: *syn* and *anti*. The most common is the *syn* orientation. An exception is Z-DNA, where only pyrimidines are found as *syn* conformers but purines are in *anti* orientation. The nucleobase plane is almost perpendicular to the plane of the sugar.

Another important conformational description is the sugar pucker. In general, a five-membered furanose is nonplanar. It can be puckered in an envelope form, which has four atoms in a plane and the fifth atom out of plane, or in a twist form with a plane through three of the atoms and two adjacent atoms on opposite sides of the plane. Ribose is normally in a form between envelope and twist but more towards the twist form. If the $C^{2'}$ or $C^{3'}$ sticks out of the plane on the same side like the $C^{5'}$, it is referred to as endo. The other side of the plane is called exo. Only two conformations are found in nucleosides: the $C^{2'}$-endo, also called south (**S**) and the $C^{3'}$-endo, also called **N** for north (*Figure 1.12*, right). A-DNA only contains $C^{3'}$-endo riboses while B-DNA occurs as $C^{2'}$-endo ribose. In Z-DNA, both forms are found: $C^{2'}$-endo for pyrimidine nucleotides and $C^{3'}$-endo for purine nucleotides.

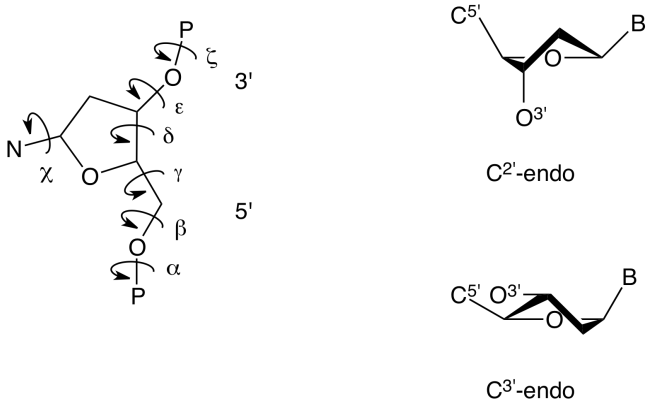


***Figure 1.12**: Torsion angles of a nucleotide unit (left) and the preferred endo sugar puckers (right).*

1.2.4.2 Secondary Structure

Three different types of DNA are known: A-DNA, B-DNA and Z-DNA. The X-ray diffraction pattern that *Watson* and *Crick* used in 1953 to deduce the helical structure came from a B-DNA.[52] The A-DNA was discovered in the same year by *Franklin* and *Gosling*.[66] They

also found that B-DNA undergoes a conformational change to A-DNA at a relative humidity below 75 % and that this process is reversible. In 1979, *Rich* and co-workers discovered a left-handed form of DNA, the Z-DNA.[67] *Table 1.4* summarises the structural characteristics of A-, B- and Z-DNA.

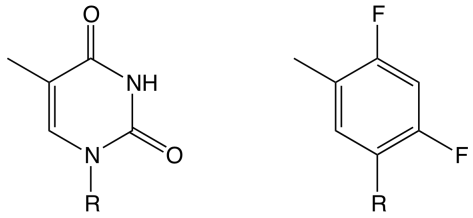
	A-DNA	B-DNA	Z-DNA
Helical sense	Right-handed	Right-handed	Left-handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11.6	10	12 (6 dimers)
Helical twist per base pair	31°	36°	9° (pyrimidine-purine); 51° (pur-pyr)
Helix pitch (rise per turn)	34 Å	34 Å	44 Å
Helix rise per base pair	2.9 Å	3.4 Å	7.4 Å per dimer
Base tilt normal to helix	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	$C^{3'}$-endo	$C^{2'}$-endo	$C^{2'}$-endo for pyrimidines; $C^{3'}$-endo for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines; syn for purines

***Table 1.4**: Characteristics of A-, B- and Z-DNA.*

The main factor contributing to duplex stabilisation is stacking. In a nucleic acid strand in aqueous solution, bases pile up in stacks similar to coins in a roll. The bases are oriented in a way such that one base plane is parallel to the adjacent base. Responsible for the stacking of nucleobases are dipole-induced dipole interactions, where a permanent dipole, mainly amino- and keto-groups, of one base are superimposed over the π-electronic system of the adjacent base.[68] Beside dipole-induced dipole interactions, also London dispersion forces[69] and hydrophobic forces[70] are responsible for base stacking.

Hydrogen bonds do not play a significant role in duplex stabilisation because of competition of the hydrogen bonds between nucleobases with water. For a long time it was assumed that hydrogen bonding only had a small contribution to duplex stabilisation but was essential for specific base pairing and to ensure the fidelity of DNA replication.[55] This changed with the surprising results in the study by *Kool* and co-workers.[71] In this famous work, the authors used 2,4-difluorotoluene (F) as a thymine isostere in a DNA replication experiment (*Figure 1.13*). Fluorine is no or only a very poor hydrogen bond acceptor, therefore

F is, unlike T, not able to form hydrogen bonds. Nevertheless, when F was incorporated in a DNA strand, A was efficiently inserted opposite F during replication with almost the same efficiency as opposite T. From this work it can be concluded that shape and steric fit are the more important factors in the selection of bases during the replication process, and not the ability to form hydrogen bonds.



T **F**

***Figure 1.13**: Structure of thymidine (T) and 2,4-difluorotoluene (F).*

1.2.4.3 Hydrogen Bonding Motifs

Under the assumption that at least two hydrogen bonds must be formed in order to obtain a stable base-pair, the four bases, when substituted at the glycosyl nitrogen (N1 and N9 for pyrimidines and purines, respectively), can be arranged in 28 different ways.[55] Two of them are the *Watson-Crick* base pairs, the other 26 are mismatches. The most important among them are shown in *Figure 1.14*.

One arrangement is the *Wobble* configuration, discovered by *Crick* in 1966.[72] The thermodynamic stability of the G-U *Wobble* base pair is comparable to the *Watson-Crick* base pairs. Furthermore, it is almost isomorphic to them. It can be found in nearly every class of RNA.[73]

A-T base pairs in solution form *Hoogsteen* base pairs. They are not found in double helices, though. These base pairs were discovered by *Hoogsteen* in 1963 and can be found in nature in tRNA, in triple helices or in G-tetrades, the sub-units of G-quadruplexes (*Figure 1.15*).[74,75]

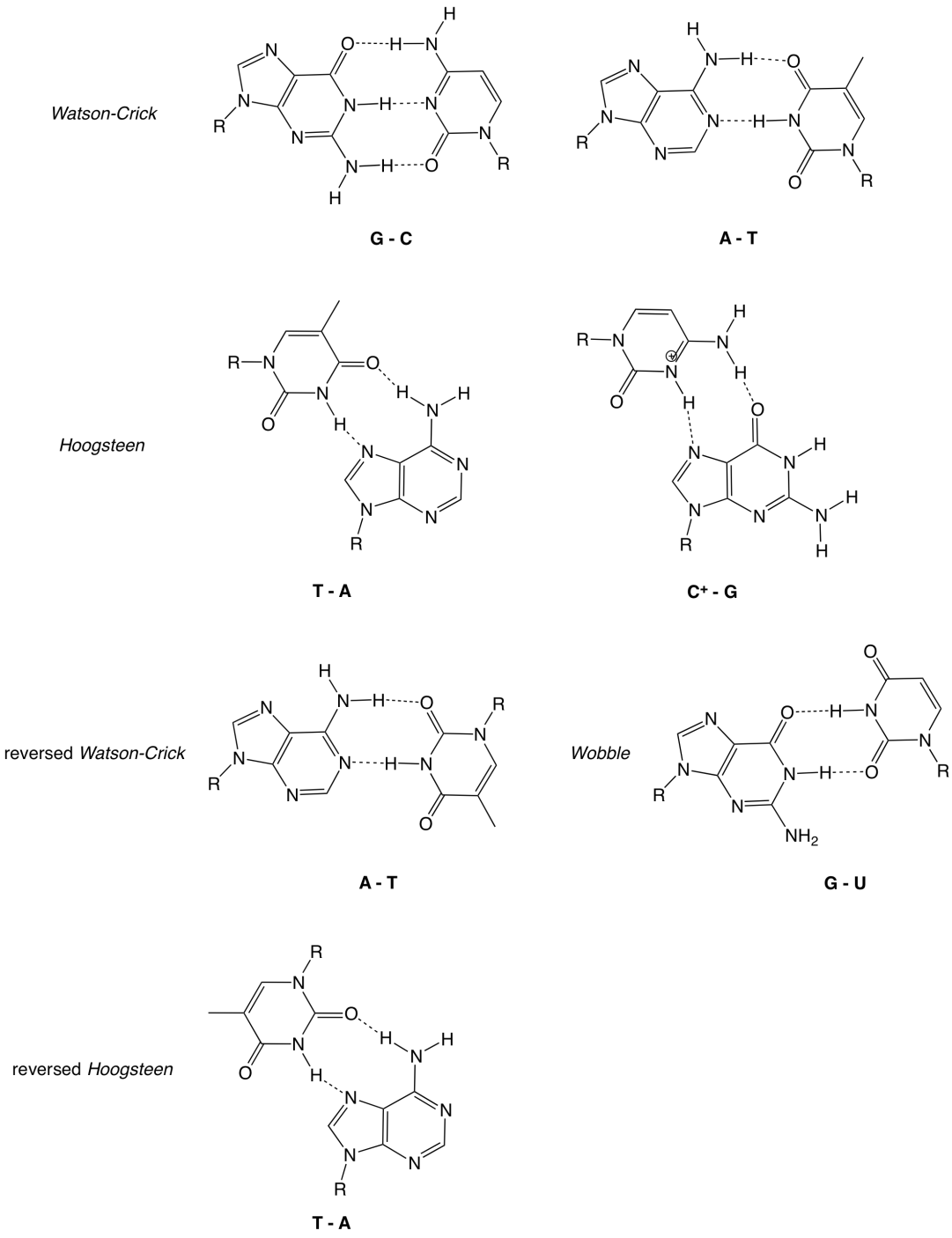


***Figure 1.14**: Hydrogen bonding motifs.*

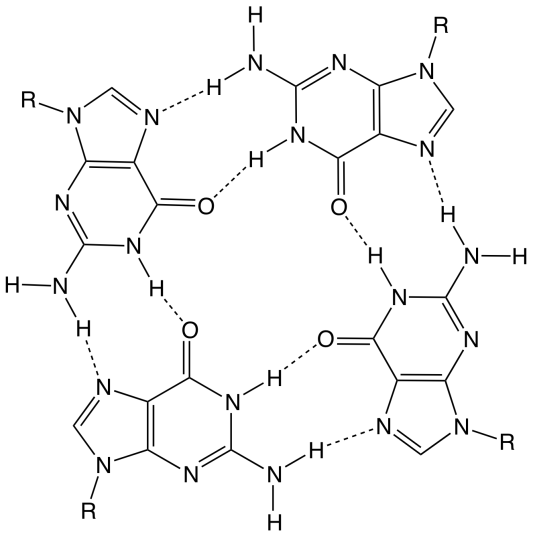
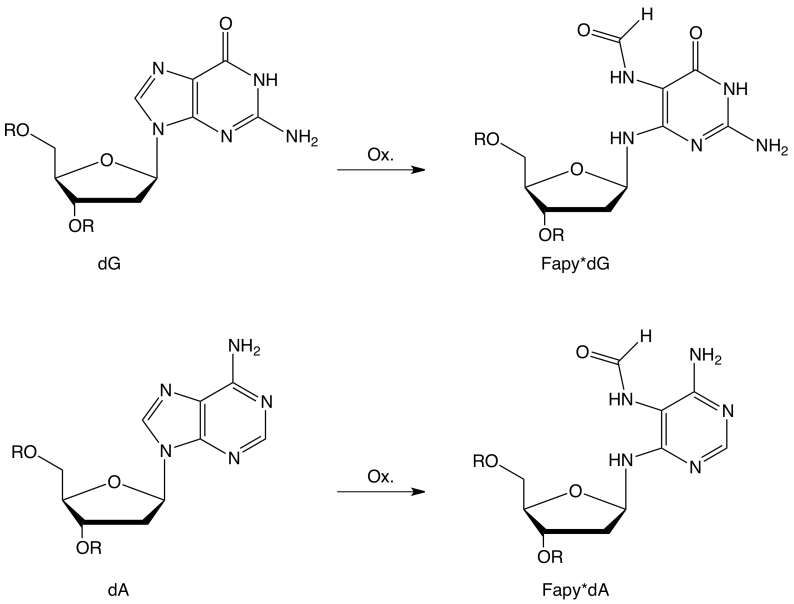


***Figure 1.15**: G-tetrade, consisting of four guanosines.*

1.3 Exocyclic Amino Nucleosides (EANs)

Several examples of natural or synthetic exocylic amino nucleosides (EANs) are known. Formamidopyrimidines can be found in nature. They result from oxidative damage to purine nucleosides and are involved in mutagenesis.[76] *Greenberg* and co-workers studied Fapy·dA and Fapy·dG, which are degradation products of dA and dG, respectively (*Scheme 1.3*).[77]



***Scheme 1.3**: Formation of Fapy·dG and Fapy·dA.*

Another important example of EANs is clitocine (*Figure 1.16*, left). It is a secondary fungal metabolite and acts as an adenosine mimics that shows cytostatic effects towards several leukaemia cell lines as well as insecticidal activities.[78,79] An example of synthetic EANs found in literature are triazine EANs (*Figure 1.16*, right).[80] They will be further discussed in the next section in terms of stability.

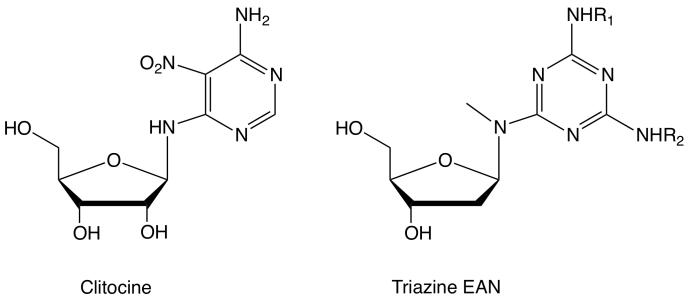
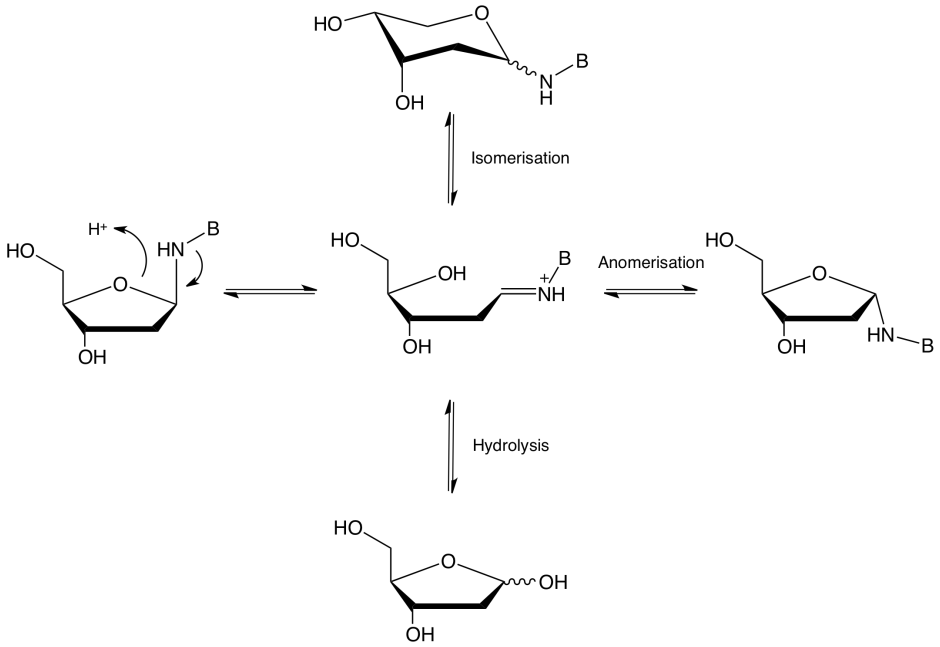


***Figure 1.16**: Molecular structure of clitocine and triazine EANs.*

1.3.1 Stability of EANs

EANs can undergo acid-induced anomerisation, isomerisation and hydrolysis (cleavage of the nucleobase), as can be seen in *Scheme 1.4*.



***Scheme 1.4**: Mechanism of Isomerisation, Anomerisation and Hydrolysis.*

Hysell et al. studied several triazine EANs and found that triazine EANs, which contain electron rich aromatic rings, are more susceptible for isomerisation and hydrolysis than electron poor ones. Also, increasing pH and protic solvents increase the rate for all decomposition processes.[80]

Especially when working with EAN systems affording nucleobases with intermediate or high electron density, a way has to be found to avoid anomerisation, isomerisation and cleavage of the nucleobase. *Greenberg* and co-workers used the fact that EANs are more stable as dinucleotides than as mononucleotides. Therefore, they synthesised a dinucleotide-phosphoramidite as starting point for the DNA chain growth. With their approach, *Greenberg* and his team were only able to incorporate an anomeric mixtures of Fapy·dG and Fapy·dA but no anomeric pure compounds.[81]

Another approach has also been described by *Greenberg* and co-workers. In this approach, the exocylic nitrogen that forms the glycosidic bond is exchanged with a carbon.[81] This increases the stability of the glycosidic bond and allows for an insertion of a defined anomer in pure form into an oligonucleotide strand. *Greenberg* and his group could show by molecular modelling[82] as well as experimentally[81] by duplex melting studies that homo-*C* analogues, modelled on Fapy·dG and Fapy·dA, are suitable models for the formamidopyrimidine lesion. Generally, a *C*-nucleoside is defined as a nucleoside with a C-C glycosidic bond instead of the naturally occurring C-N glycosidic bond. A subclass of *C*-nucleosides are homo-*C*-nucleosides. Their sugar moiety is connected by a methylene bridge to a carbon atom of the nucleobase.[83]

A third approach by *Carell* and team strengthens the glycosidic bond by replacement of the furanose sugar moiety by cyclopentane (*Figure 1.17*). This approach also allows for incorporation of anomeric pure EANs into an oligonucleotide strand.[84]

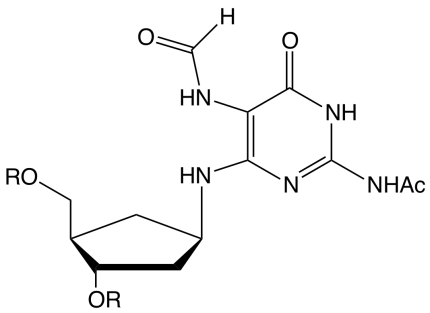


***Figure 1.17**: Fapy·dG with a cyclopentane- instead of a sugar-moiety.*

1.4 Synthesis of Oligonucleotides

The first chemical synthesis of a dinucleotide with a 3' → 5' internucleotidic linkage, which was identical to natural DNA, was reported in 1955 by *Michelson* and *Todd*.[85] In their pioneering approach, they activated the 5'-protected thymidine-3'-*O*-benzyl hydrogen phosphonate by *N*-chlorosuccnimide (NCS) and added a 3'-protected thymidine. After removal of the protecting groups, the natural thymidylic acid was obtained.

The first oligonucleotide synthesis that used a solid support was investigated by *Letsinger* and *Mahadevan* in the mid-sixties.[86] The major advantage of this concept, that was already known from peptide synthesis, is the facile separation of the starting material and reagents in the soluble phase from the product, which is covalently bound to the solid support. Thus, no time-consuming purification steps were needed. Nowadays, the most commonly used approach is the phosphoramidite approach, developed by *Beaucage* and *Caruthers*.[87] The reaction cycle of the phosphoramidite methodology is shown in *Figure 1.18*.

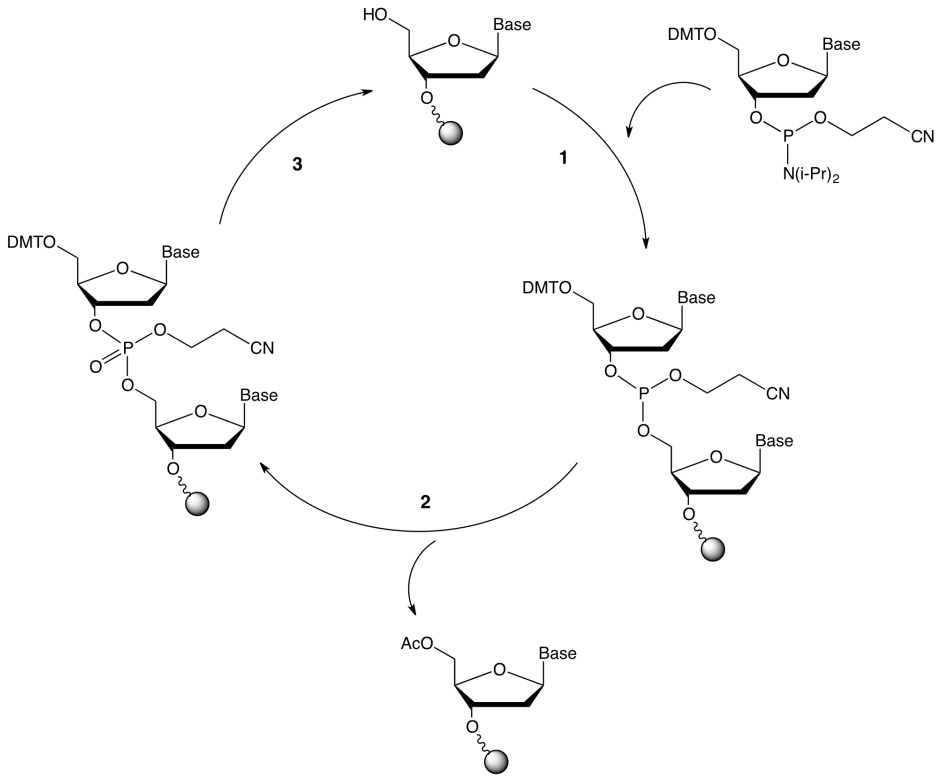
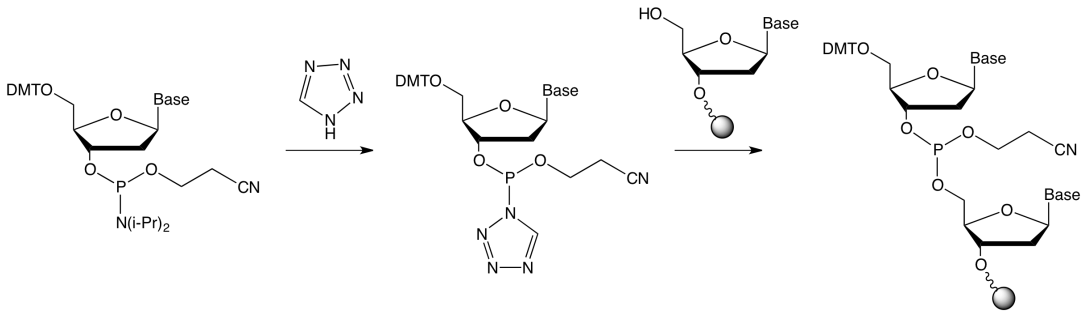


***Figure 1.18**: Synthetic cycle of the phosphoramidite methodology.*

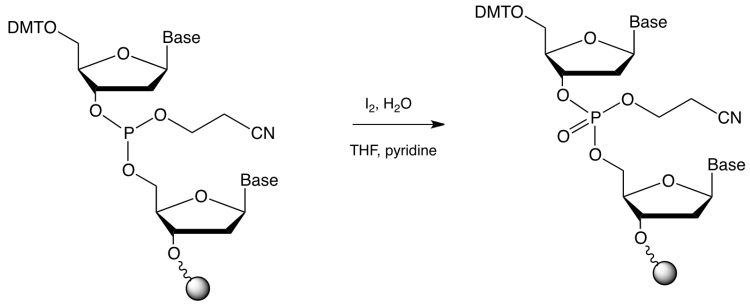
The phosphoramidite methodology follows a 5' → 3' chain elongation direction. The cycle contains three steps: the coupling step, a capping/oxidation step and the detritylation. In the end, the strand is cleaved from the solid support.

In the coupling step, the phosphoramidite is typically activated with 1*H*-tetrazole, then reacted with a 5'-deprotected nucleoside bound to the solid support (*Scheme 1.5*).



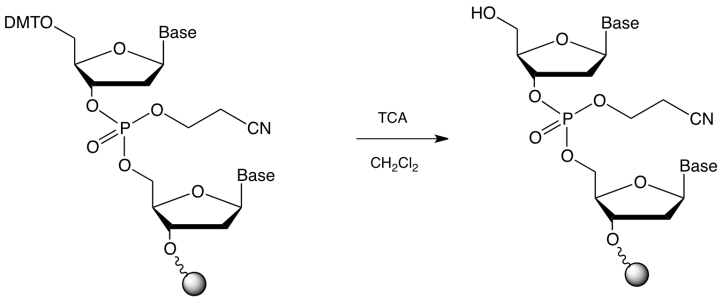
***Scheme 1.5**: Phosporamidite activation and coupling step.*

The next step includes capping and oxidation. There are always some nucleosides, that will not succeeded in coupling in the previous step and therefore still contain a free 5'-OH group. These uncoupled hydroxy groups are treated with a mixture of 1-methylimidazole in THF and 2,6-dimethylpyrimidine in Ac_2O in order to deactivate them for further reactions in the following steps. After capping, the phosphite is oxidised by treatment with I_2 and water in THF/pyridine to the phosphate (*Scheme 1.6*).



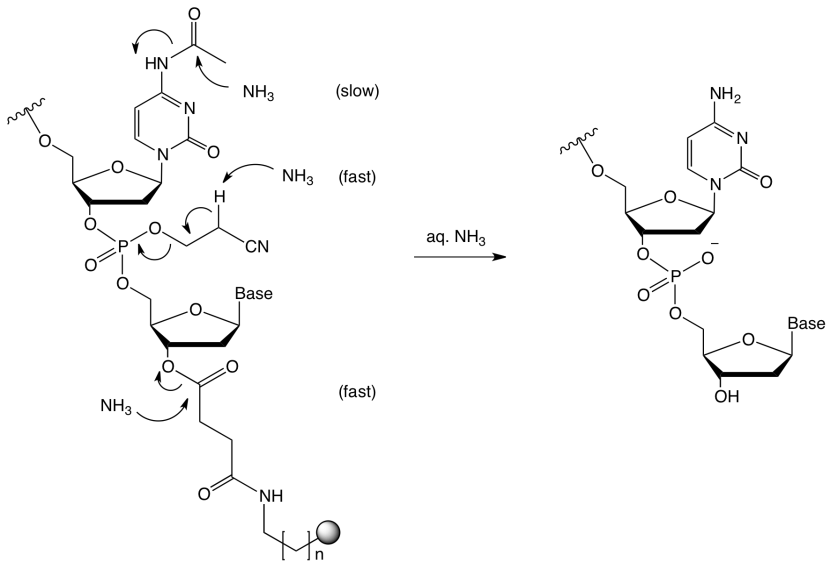
***Scheme 1.6**: Oxidation of the phosphite.*

In the last step of the cycle, the DMT protecting group at the 5'-position is cleaved under acidic conditions, usually by treatment with trichloroacidic acid (TCA) in CH_2Cl_2. Then, the 5'-end of the strand is ready for the next coupling step (*Scheme 1.7*).



***Scheme 1.7**: Deprotection step.*

After the last cycle, the oligonucleotide is treated with aq. NH_3-solution at elevated temperature in order to fully deprotect the oligonucleotide strand and remove it from the solid support. Under basic conditions, all protecting groups on the nucleobases and on the phosphates are cleaved and the linker to the solid support is released (*Scheme 1.8*).



***Scheme 1.8**: Cleavage of protecting groups and solid support.*

1.5 ITC

To determine the binding strength and stability of a duplex, the most commonly used technique is the recording of melting curves by UV or circular dichroism (CD). Thereby, a single binding constant K_d is determined, from which $\Delta G°$ can be calculated by *Equation 1.1*. To determine the thermodynamic parameters $\Delta H°$ and $\Delta S°$, a *van't Hoff* analysis is necessary.

In a *van't Hoff* analysis, K_a values at different temperatures are measured. Then, a plot of $1/T$ versus $\ln K_a$ gives the $\Delta H°$ and $\Delta S°$ values.

$$\Delta G° = -RT\ln(K_a) \quad (1.1)$$

$$\Delta G° = \Delta H° - T\Delta S° \quad (1.2)$$

However, isothermal titration calorimetry (ITC) gives K_a and $\Delta H°$ directly in one analysis. By using *Equations 1.1* and *1.2*, $\Delta G°$ and $\Delta S°$ can easily be calculated. Upon association of a host (H) and a guest (G), heat will either be released or absorbed. In ITC, the heat released or absorbed in each titration step is measured by comparison of the temperature changes that occur between a sample cell where guest is added to host, and a reference cell with the same medium as in the sample cell but lacking the guest and host (shown in *Figure 1.19*).[88]

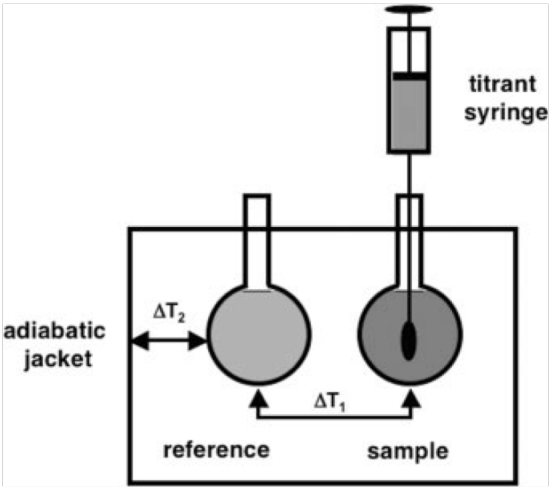


***Figure 1.19**: Diagram of an isothermal titration calorimeter.*[89]

Equation 1.3 shows the total heat (Q) released or absorbed for all the additions of guest, where V is the volume of the vessel. Now, the binding isotherm for [H•G] is needed (*Equation 1.4*). A binding isotherm is the theoretical change in the concentration of one component as a function of the concentration of another component at constant temperature. *Equation 1.4* is derived from *Equation 1.5*, using the relationship that $[H]_0 = [H•G] + [H]$,

where $[H]_o$ is the initial concentration of H and [H•G] is the concentration of host-guest complex.[90]

$$Q = V\Delta H^\circ\,[H\bullet G] \tag{1.3}$$

$$[H\cdot G] = \frac{[H]_o K_a [G]}{1 + K_a [G]} \tag{1.4}$$

$$K_a = \frac{[H\cdot G]}{[H][G]} \tag{1.5}$$

Using the binding isotherm (*Equation 1.4*) for [H•G] gives *Equation 1.6*.

$$Q = \frac{V\,\Delta H^\circ K_a [H]_o [G]}{1 + K_a [G]} \tag{1.6}$$

Now, why can the heat released or absorbed directly be related to the enthalpy? The first law of thermodynamics state that the internal energy of a closed system is constant, thus:

$$\mathrm{d}U = \mathrm{d}q + \mathrm{d}w \tag{1.7}$$

with U being the internal energy, q the heat and w the work. The enthalpie H is given in *Equation 1.8*.

$$H = U + p\cdot V \tag{1.8}$$

Generally, from a change of state follows $U \rightarrow U + \mathrm{d}U$, $p \rightarrow p + \mathrm{d}p$ and $V \rightarrow V + \mathrm{d}V$, thus, from *Equation 1.8*, a change in enthalpy can be written as:

$$H + \mathrm{d}H = U + \mathrm{d}U + (p + \mathrm{d}p)(V + \mathrm{d}V)$$

$$= U + \mathrm{d}U + p\cdot V + p\cdot\mathrm{d}V + \mathrm{d}p\cdot V + \mathrm{d}p\cdot\mathrm{d}V$$

$U + p\cdot V$ can be substituted with H (from *Equation 1.8*) and $\mathrm{d}p\cdot\mathrm{d}V$ is negligible because it is the product of two infinitesimal variables. With *Equation 1.7*, it follows:

$$dH = dq + dw + p \cdot dV + dp \cdot V \quad (1.9)$$

In a mechanical equilibrium with the environment, $dw = -p \cdot dV$ can be substituted in *Equation 1.9*, giving:

$$dH = dq + dp \cdot V \quad (1.10)$$

At constant pressure ($p = 0$), it follows:

$$dH = dq \quad (1.11)$$

And for measurable changes:

$$\Delta H = q_p \quad (1.12)$$

An isobaric calorimeter like ITC measures q, and so we obtain directly ΔH, the change in enthalpy.

Despite the fact that ITC can directly deliver a complete thermodynamic characterisation from one single titration, there are not many examples for measurements of DNA duplex formation by ITC.[91,92] To determine the binding strength of DNA towards another DNA strand, melting curves or differential scanning calorimetry (DSC) are most commonly used. The reason might be that the first ITC instruments needed relatively big amounts of material – a requirement that limits the usability for DNA. However, the sensitivity of these instruments has improved dramatically over the past 15-20 years.[89] Thus, sample requirements are now accessible for biological systems like DNA. One example, where the complexation between oligonucleotides was investigated by ITC, is a study by *Bruylants et al.* The authors compared the pairing stability of a locked nucleic acid (LNA)-DNA heteroduplex and of its isosequential DNA-DNA homoduplex.[91] They did not measure single nucleoside differences, though.

2 Problem Outline

2.1 The Project

Of the present day nucleobases, cytosine is the least stable and is known to undergo hydrolysis to uracil. Based on this fact, cytosine can be seen as a hydrolysis product of 2,4-diaminopyrimidine (*Figure 2.1*).

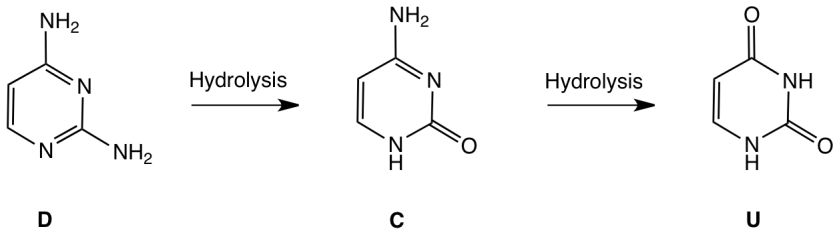


***Figure 2.1**: Hydrolysis of 2,4-diaminopyrimidine (D) to cytosine (C) and uracil (U).*

With these three pyrimidine bases as building blocks, an all-pyrimidine base pairing code can be formed, as proposed by *Siegel* and *Tor*.[39] It consists of the two natural nucleosides cytidine and uridine as well as the two unknown nucleosides D and E, both forming their glycosidic bond through an exocyclic amino group. D, containing a 2,4-diamionpyrimidine nucleobase, is isosteric to adenosine (A) and can form base pairs with U. E is based on a cytosine nucleobase but in contrast to C, its base is connected via the exocyclic amino function in the 4-position to the ribose moiety. E is isosteric to guanosine (G) and forms base pairs with C (*Figure 2.2*).

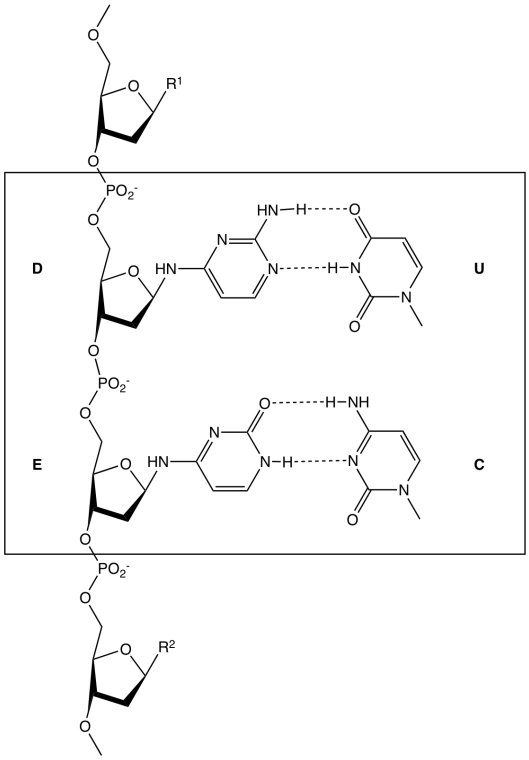


***Figure 2.2**: All-pyrimidine base pairing code.*

Following our hypothesis, EAN's D and E are harbingers of A and G, respectively. EAN's are known to be degradation products of purines in nature under oxidative stress[93] (see *section 1.3*). Assuming an oxidative atmosphere on the early earth, purines could not have existed without a repair system but EAN's could. Therefore, EAN's are plausible precursors of purine nucleosides.

A possible precusour of present-day DNA/RNA should fulfil certain criteria.[39] It should be compatible with the present system but less fit for environmental pressure to allow for mutational evolution. It should be isosteric with the modern system but thermodynamically less stable. Additionally, it should fulfil the single source criterion, meaning that a single heterocycle should have established a protocode. 2,4-diaminopyrimidine meets all these criteria. In addition, it was synthesised under presumably prebiotic conditions from guanidine hydrochloride and cyanoacetaldehyde, as shown by *Miller* and coworkers[94], and from cyanat and cyanoacethylene by *Shapiro*[95].

2.2 State of Knowledge

Despite the successful synthesis of pyrimidine based EAN's by *Greenberg et al.*[81], earlier attempts in our group to synthesise exocylic amino nucleoside dD failed. *Greenberg's* synthesis seems to be limited to electron-poor aromatic rings. Furthermore, EAN's can easily undergo anomerisation, furanose-pyranose isomerisation and cleavage of the glycosidic bond.[77,80] For these synthetic as well as stability reasons, we switched to carbon analogues, since studies by *Greenberg* and co-workers could show them to be suitable model compounds.[77,99] Additionally, *Berstis* could show computationally that 2,4-diaminopyrimidine (D-NH2) and 2-aminopyrimidine (D) afford a similar electron distribution like their natural analogue adenine (*Figure 2.3*).[96]

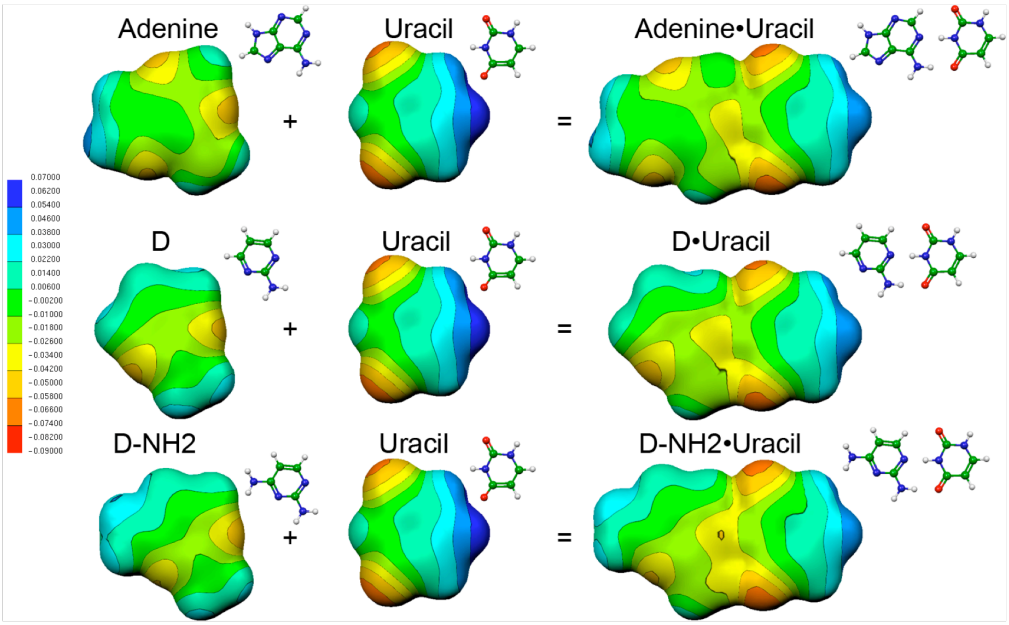
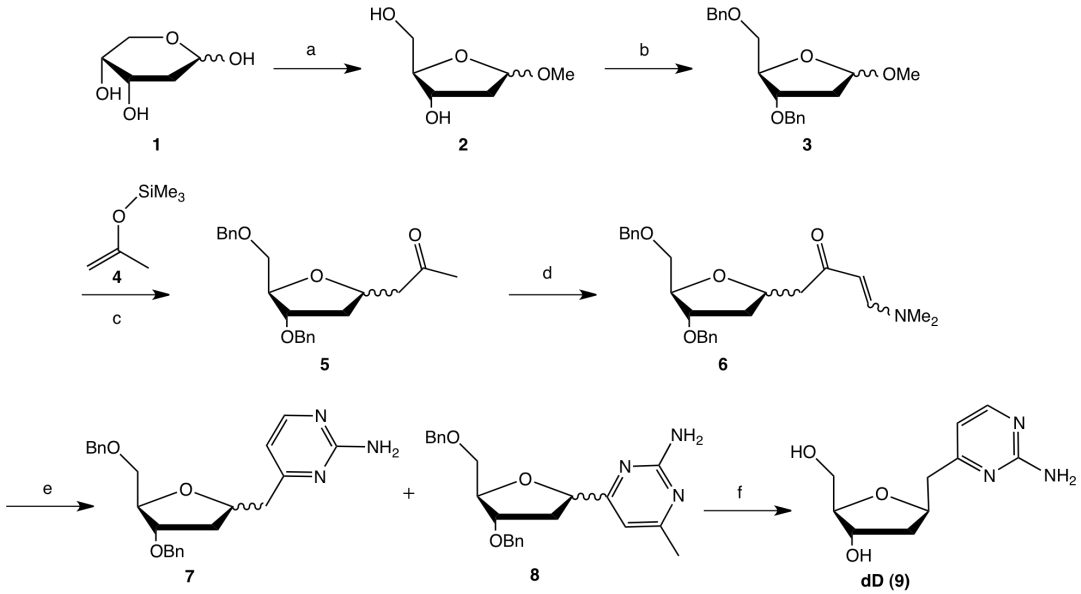


***Figure 2.3**: Calculated electron distributions for adenine, 2-aminopyrimidine (D) and 2,4-diaminopyrimidine (D-NH2).*

A strategy for the synthesis of homo-*C*-nucleoside dD was developed in our group and is shown in *Scheme 2.1*.[97,98] The synthesis starts from commercially available 2-deoxy-D-ribose and proceeds in three steps to compound **5**. By treatment of **5** with $Zn(OAc)_2$ under basic conditions, the α/β ratio can be shifted to a two-fold excess of the desired β-anomer. Unfortunately, the anomers cannot be separated by column chromatography. The synthesis is continued with the addition of the *Bredereck's* reagent, followed by condensation of guani-

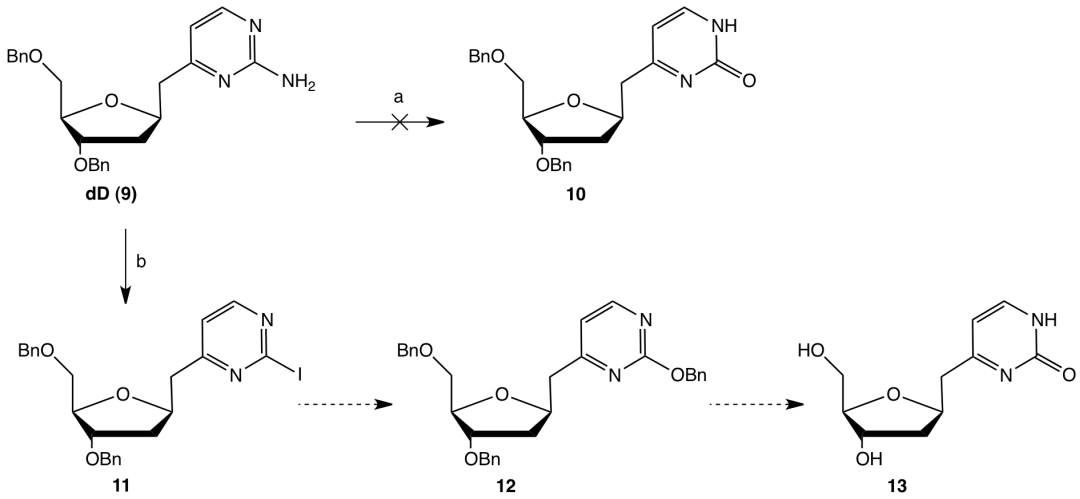
dinium sulfate. This leads to aminopyrimidine **7** as main product but also to side product **8**, which results from the addition of the *Bredereck's* reagent at the position C(1) instead of the C(3) of **5**. By tuning the reaction conditions, the amount of side product could be reduced but not completely avoided. Furthermore, of the main product as well as the side product, an α- and β-anomer are obtained. This leads to four different products, which are not separable by column chromatography but need to be separated by HPLC. Finally, dD (**9**) can be obtained by deprotection of the isolated β-anomer of **7** with BBr_3.



a) MeOH, HCl, rt, 15 min, 98 %. b) NaH, $INBu_4$, BnBr, THF, rt, 12 h, 58 %. c) i) trimethyl(prop-1-en-2-yloxy)silane **4**, $SnCl_4$, MeCN, 0 ° C, 2h, 60 %; ii) $Zn(OAc)_2$, NaOMe, MeOH, rt, 3 d, 99 %. d) $tBuOCH(NMe_2)_2$, Tol, 70 °C, 20 h. e) i) NaOEt, EtOH, 60 °C, 4 h; ii) $(CH_3N_3)_2*H_2SO_4$, 40 °C to 70 °C, 18 h, 22 % over two steps. f) BBr_3, Ch_2Cl_2, 0 °C, 3 min, 99 %.

Scheme 2.1: *Synthesis of dD.*

In the same work, several attempts to synthesise homo-*C*-nucleoside dE were undertaken (*Scheme 2.2*.).[97]



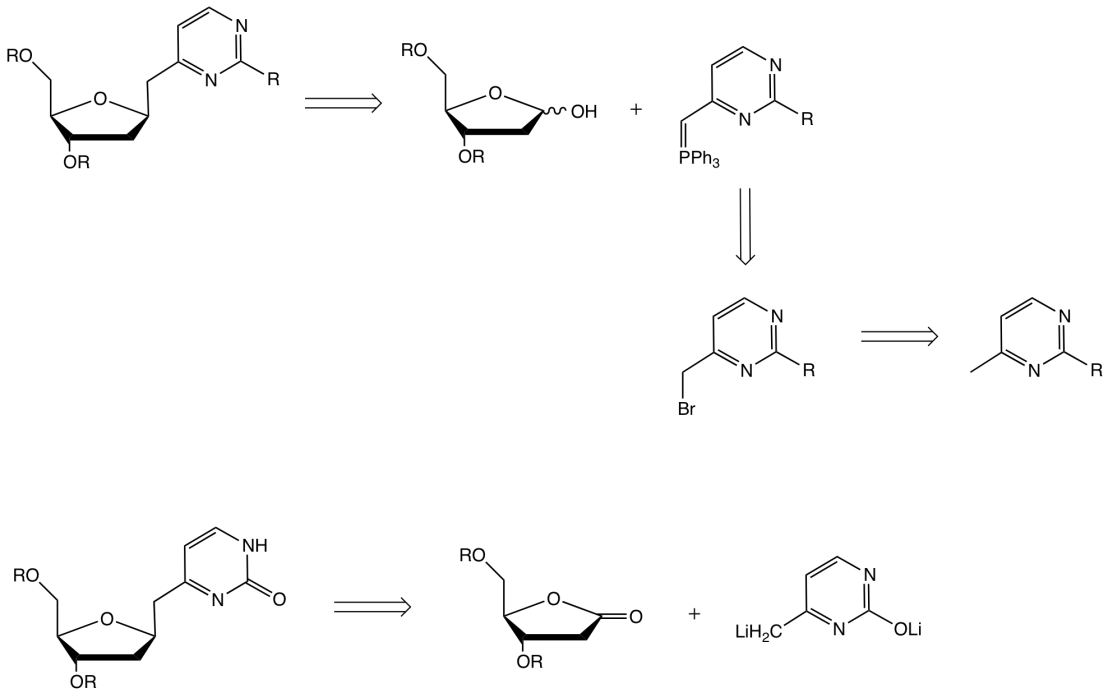
a) 10 M NaOH, H_2O, 90 °C, 2 d. b) $(C_5H_{11})ONO$, CH_2I_2, CuI, THF, 60 °C, 1.5 h, 19 %.

***Scheme 2.2**: Attempted synthesis of dE.*

The first attempts focused on synthesising dE from dD. Simple hydrolysis of dD to **10** under basic conditions was not successful. Therefore, the conversion of dD to its 2-iodopyrimidine **11** was tested but after screening various reaction conditions, no yield higher than 19 % could be obtained.

Because of these low yields, alternative strategies were tested. *Scheme 2.3* shows two different strategies. The first one pursued a Wittig reaction but the Wittig reagent could not be synthesised. Treatment of different 4-methylpyrimidine compounds with NBS always resulted in 5-bromo-4-methylpyrimidine instead of the desired bromination of the methyl group. The other strategy shown in Scheme 2.3 sought a coupling of the lactone of deoxyribose to a double lithiated 4-methylpyrimidone. This strategy was not successful due to a ring opening during the nucleophilic attack.

Despite all these different synthesis attempts it was not possible to obtain dE.



***Scheme 2.3**: Alternative strategies.*

3 Results and Discussion

3.1 Retrosynthetic Analysis

The existing synthesis of dD is described in *Section 2.2*. This synthesis works but suffers from several drawbacks. The main disadvantage is the need for an HPLC separation, which is very time-consuming and costly and does not allow the synthesis of bigger quantities of material. Additionally, the yield of the key step, the addition of the *Bredereck's* reagent followed by guanidinium condensation, at 22 %, is rather low. Furthermore, it was not possible to obtain dE via this synthetic route. For all these reasons, we decided to investigate different synthesis ways that are more time efficient and allow the production of bigger amounts of material.

Conceptually, the retrosynthetic disconnection can be made between the C(1') of the ribose and the exocyclic methylene bridge (*Figure 3.1*, disconnection a) or between the bridging CH_2 and the C(4) of the pyrimidine ring (*Figure 3.1*, disconnection b).

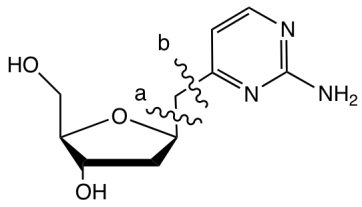
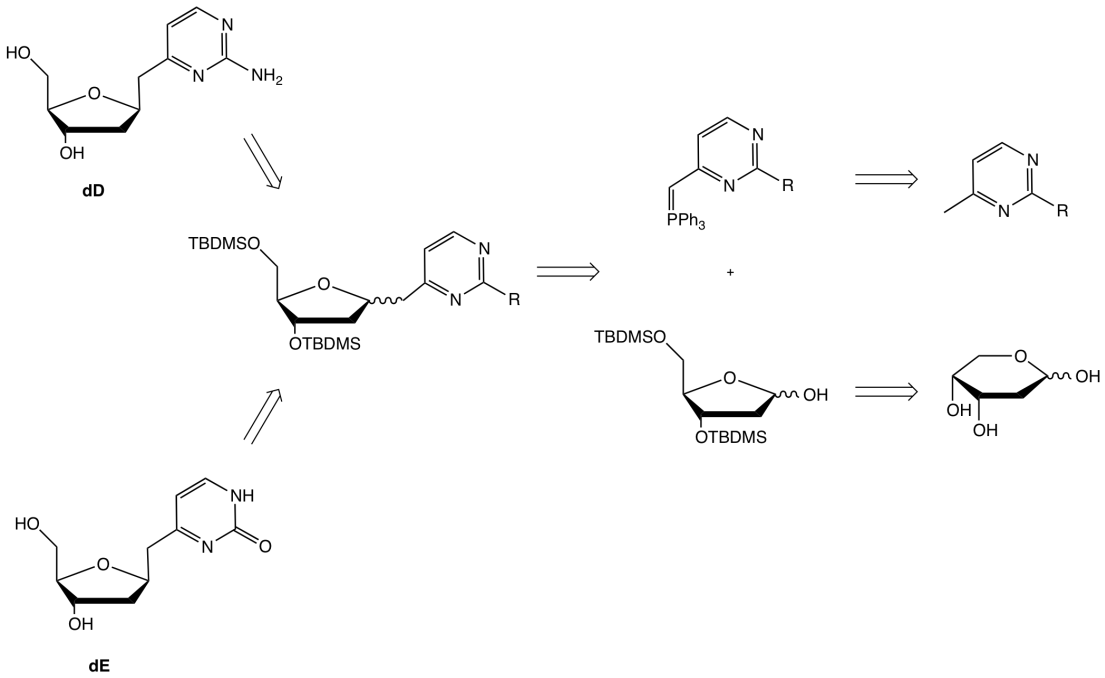


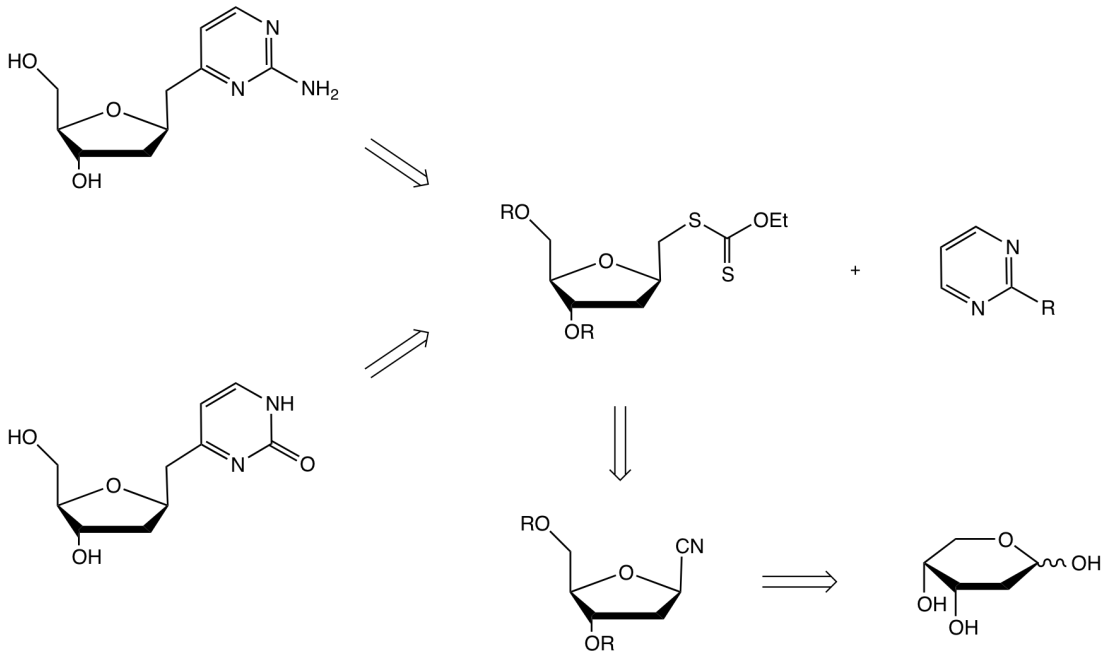
***Figure 3.1**: Possible disconnection of dD.*

The first synthetic approach towards dD and dE uses a disconnection next to the C(1') of the deoxyribose moiety. The sugar ring is coupled with the pyrimidine ring by a Wittig reaction (*Scheme 3.1*). The R-group on the pyrimidine ring can then be modified to an amino group and oxygen in order to obtain dD and dE, respectively. This approach will be explained in more detail in *Section 3.2.2*.



Scheme 3.1*: Retrosynthetic Wittig approach for the synthesis of dD and dE.*

Further retrosynthetic analysis revealed the route shown in *Scheme 3.2*. In this synthetic pathway, the disconnection is made next to the pyrimidine ring. This has the advantage that the configuration of the compound is already fixed before the coupling of the two building blocks. At this stage, the α- and β-anomer can still be separated by column chromatography. When using diastereopure starting material, this would offer a possibility to obviate the time-intense and costly HPLC separation. A detailed description of the radical-type pathway can be found in *Section 3.2.2*.



Scheme 3.2: *Retrosynthetic radical-type pathway for the synthesis of dD and dE.*

3.2 Synthesis

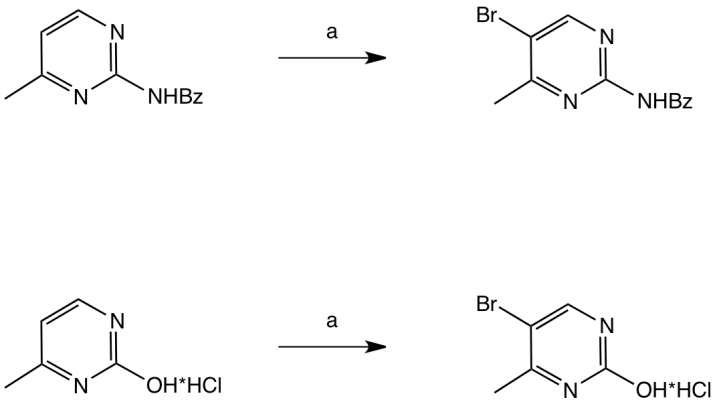
In literature, a variety of *C*-nucleosides have been reported but only few of them contain a methylene bridge connecting the sugar moiety with the pyrimidine ring (so called homo-*C*-nucleosides; for definition of these terms, see *Section 1.3.1*).[99,100]

Here, two different synthetic routes to dD and dE are described. First, a pathway using a *Wittig* reaction to couple the sugar ring to the pyrimidine nucleobase is discussed in *Section 3.2.1*. Second, an attempt including a radical reaction in the coupling step is presented in *Section 3.2.2*.

3.2.1 Wittig Approach

A *Wittig* reaction, discovered by *Georg Wittig* in 1954, is the reaction of an aldehyde or ketone and a phosphonium ylide, usually a triphosphonium ylide, to form an alkene.[101] Triphosphonium ylide is also called *Wittig* reagent. Its synthesis proceeds through a bromination, followed by conversion to the triphenylphosphonium salt and reduction to the ylide. Earlier attempts in our group to synthesise the *Wittig* reagent of the pyrimidine bases (shown

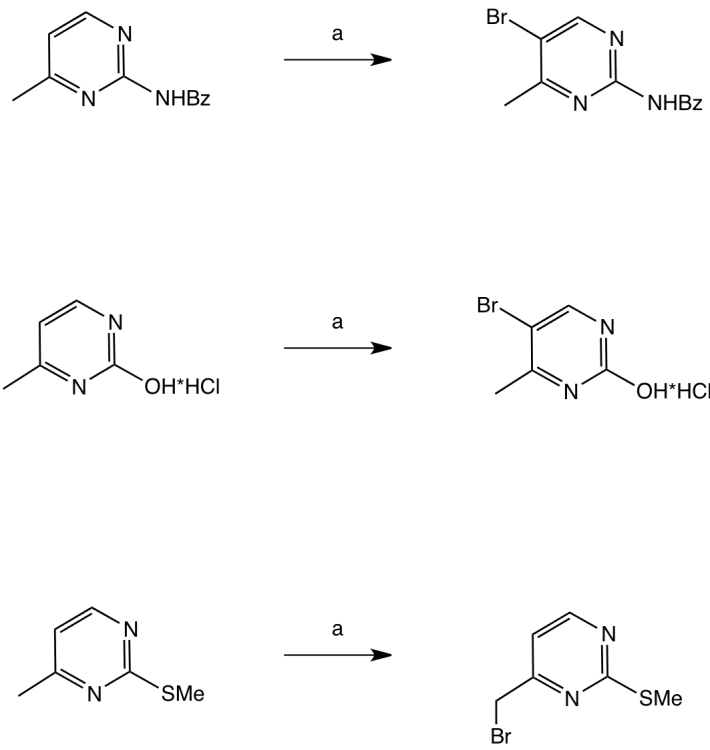
in *Scheme 3.3*) failed, since bromination by NBS always occurred in the 5-position directly on the ring instead of the desired bromination of the exocyclic methyl group on the C(4).[97]



a) NBS, DBP, CH_3Cl, 2 h, 55 °C.

***Scheme 3.3**: Bromination by NBS in the 5-position of the pyrimidine ring.*

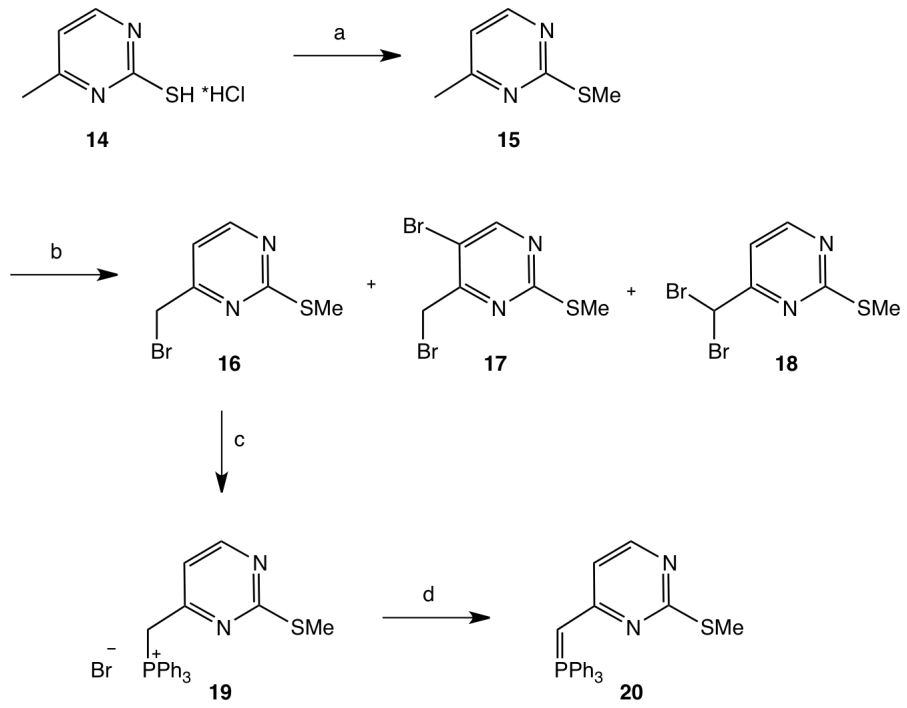
On the other hand, *Strekowski et al.* were able to brominate 4-methyl-2-(methylthio)pyrimidine on the methyl group by using Br_2 in acetic acid.[102] Following these results, we treated 2-hydroxy-4-methylpyrimidine hydrochloride with Br_2 in acetic acid but obtained bromination in the 5-position (*Scheme 3.4*). *Steinauer* observed the same 5-position bromination using benzoyl protected 2-amino-4-methylpyrimidine.[103]



a) Br_2, AcOH, 80 °C, 2 h.

***Scheme 3.4**: Bromination by Br_2.*

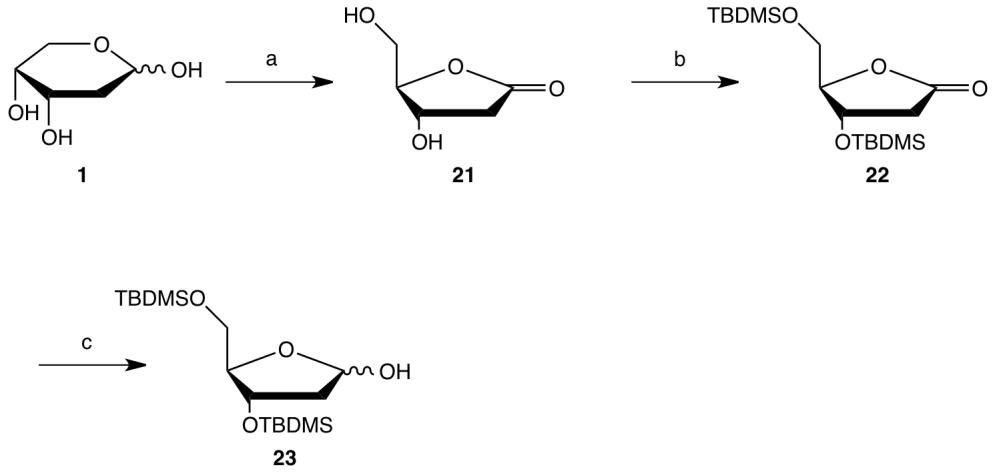
On account of these results we decided to use the thiomethylpyrimidine to form the *Wittig* reagent and, in a later step, to modify the thiogroup to the desired amino or keto group. This procedure has the additional advantage that only one precursor is needed to obtain both, dD and dE. In *Scheme 3.5*, the synthesis of the *Wittig* reagent is shown. It starts from commercially available 4-methylpyrimidine-2-thiol hydrochloride. Methylation by iodomethane afforded 4-methyl-2-(methylthio)pyrimidine **15** in good yield.[104] The following bromination step yielded the desired product **16** brominated on the exocyclic methyl group as main product in 54 % yield.[102] Two different dibromo sideproducts, **17** and **18**, were observed in small quantities (5 % and 3 %, respectively). Also a considerable amount of starting material could be recovered. The product, as well as the two side products, were obtained as lightly coloured oils that decompose overnight at room temperature to black sticky solids. The product should therefore be used immediately in the next reaction step or stored at -20 °C for not longer than a day. Bromopyrimidine **16** was then treated with triphenylphosphine to give the phosphonium salt **19**, followed by elimination to the Wittig reagent **20**.[100]



a) MeI, NaOH, H_2O, EtOH, rt, 4 h, 82 %. b) Br_2, AcOH, 80 °C, 2 h, 54 %. c) PPh_3, benzene, rt, 4 h, 60 %. d) NaOH, H_2O, H_2Cl_2, rt, 1 h, 82 %.

***Scheme 3.5**: Synthesis of the Wittig reagent.*

The synthesis of the second building block, the deoxyribose moiety, is shown in *Scheme 3.6.*

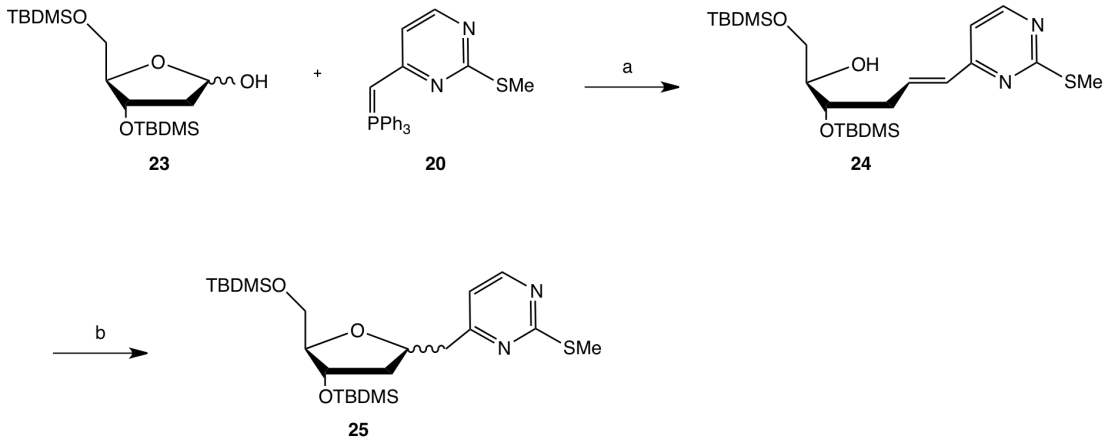


a) Br_2, H_2O, 0 °C, 16 h, 99 %. b) TBDMSCl, DMF, rt, 21 h, 63 %. c) DIBAlH, -78 °C, 5 h, 95 %.

***Scheme 3.6**: Synthesis of the deoxyribose building block.*

Commercially available 2-deoxy-D-ribose was oxidised to the lactone **21** by bromine in water in quantitative yield.[105] Next, the hydroxy groups were protected by TBDMS protecting groups. Finally, the compound was reduced by DIBAlH to obtain lactole **23** in very good yield, according to a literature procedure.[106]

With these two building blocks in hand, a *Wittig* reaction could be performed to link the sugar moiety to the pyrimidine nucleobase (*Scheme 3.7*). Lactole **23** is in equilibrium with its open ring aldose form, which provides the aldehyde needed for the *Wittig* reaction to proceed. This aldehyde reacted with the phoshponium ylide to form **24** and triphenylphosphine oxide. Intermediate **24** was then treated with NaOMe in MeOH to obtain **25** in an anomeric mixture. Traces of **24** could be separated by column chromatography and again subjected to NaOMe in MeOH to repeat the ring closing step and obtain more product **25**.



a) i) Toluene, 120 °C, 3 d; ii) NaOMe, MeOH, rt, 1 h, 72 %.

***Scheme 3.7**: Wittig reaction.*

The α- and β-anomer of **25** could not be separated by column chromatography, thus, HPLC was used for the separation. First, several attempts were made in reverse phase with various water-acetonitrile gradients on a C5, a phenyl and different C18 columns. All of these attempts showed only one peak and no separation. Then, it was switched to a normal phase system where we obtained slight separation on an amino column and finally satisfying separation on an nitrile column with a isocratic hexane + 0.1 % isopropanol eluent (*Figure 3.2*).

Using this eluent, we were able to separate a sufficient amount of the anomeric mixture to obtain the pure anomers.

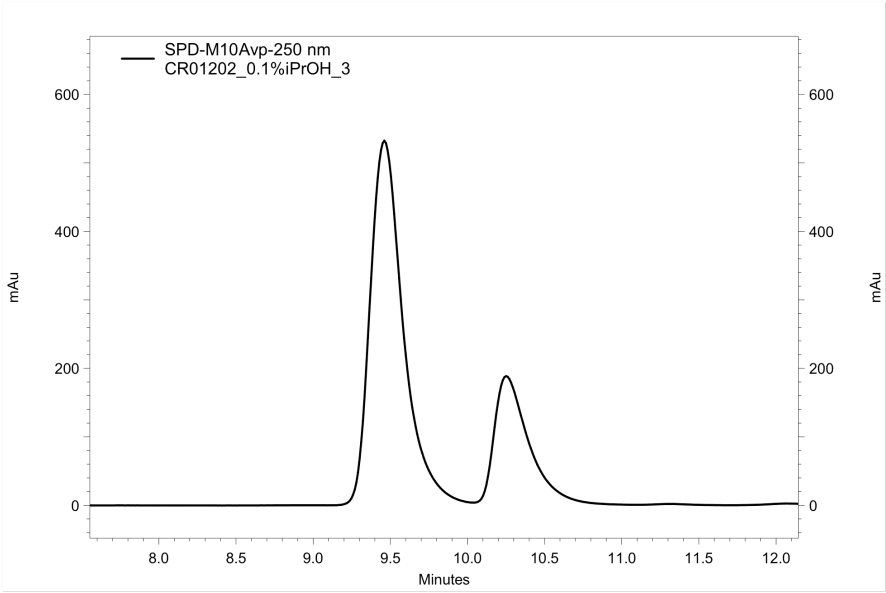


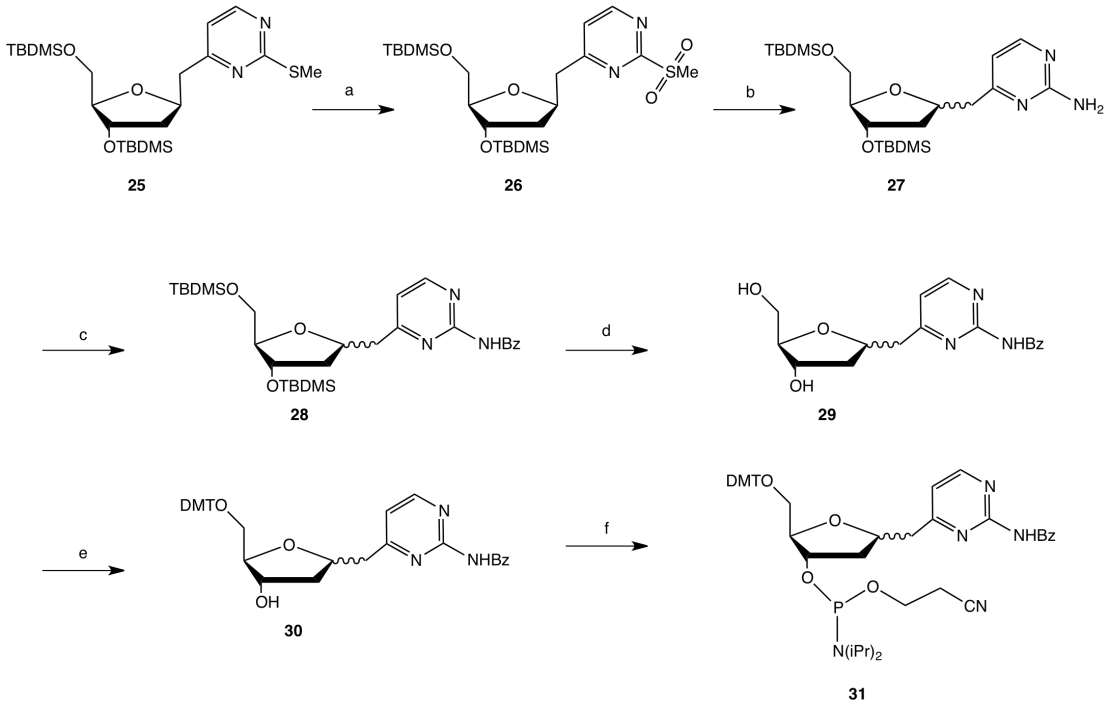
***Figure 3.2**: HPLC separation of α- and β-anomer of **25** with hexane + 0.1 % iPrOH as eluent.*

2D-NMR measurements (COSY, NOESY) were performed on both anomers separately in order to find out which is the desired β-anomer. With help of these measurements it was possible to assign the bigger peak in the HPLC absorbance spectrum to belong to the β-anomer and the second, smaller peak to the α-anomer. Thus, the *Wittig* reaction resulted in a **25α**/**25β**-ratio of 1:2.

Pure compound **25β** was then used as a precursor for the synthesis dD, as described below.

3.2.1.1 Synthesis of dD and its CE PA

The synthesis towards dD and its cyanoethyl phosphoramidite (CE PA) was continued as shown in *Scheme 3.8*.



a) mCPBA, CH_2Cl_2, rt, overnight, 78 %. b) aq. NH_3, dioxane, 80 °C, 4 h, 77 %. c) i) BzCl, pyridine, rt, 2 h; ii) aq. NH_3, 0 °C, 30 min, 73 %. d) NH_4F, MeOH, 70 °C, 15 h, 95 %. e) DMTCl, DMAP, pyridine, CH_2Cl_2, rt, overnight, 80 %. f) $PCl(N(iPr)_2OEtCN)$, $NEt(iPr)_2$, CH_2Cl_2, rt, 2 h, 83 %.

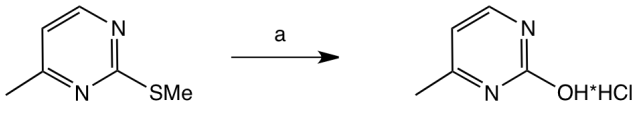
***Scheme 3.8**: Synthesis of dD CE PA (**31**).*

The first step was the oxidation of the thiol **25β** to sulfone **26β** by treatment with mCPBA.[107] In the next step, sulfone **26** was reacted with aq. NH_3 in dioxane at 80 °C to give amine **27**, according to a literature procedure.[108] Despite the fact, that only the β-anomer was used as educt in this reaction, an anomeric mixture of α/β 1:5 was obtained of amine **27**. Anomerisation seemed to occur during the reaction in the presence of base at elevated temperatures. Using the same reaction conditions but increasing the temperature to only 40 °C did not show any improvement. The reaction did not proceed completely and anomerisation could still be observed. These results were unsatisfying since only the pure β-anomer is of interest for this project and the result of the labour intense HPLC purification step was destroyed during this amination step. The synthesis was continued with the α/β mixture to test the feasibility of the synthesis and also to see if addition or substitution of protecting groups have an influence on the separation behaviour of the anomers by column chromatography. For example, *Delaney* and *Greenberg*[99] could separate the α- from the β-anomer of their homo-*C*-nucleoside when protected with a trityl group on the 5'-hydroxy group.

In the next step, the amino function was protected with a benzoyl protecting group. Following a literature procedure[109], selective monobenzoylation was achieved by treatment with aq. ammonia after the addition of benzoyl chloride. It followed deprotection with NH_4F as fluoride source to yield diol **29**.[110] The final steps, following well-known literature procedures, include protection of the 5' hydroxy group by the bulky dimethoxytrityl protecting group[111] (which selectively adds to the primary 5'-hydroxy group but not the secondary 3'-hydroxy group due to steric hindrance) followed by a conversion of the 3'-hydroxy group to a cyanoethyl-*N*,*N*-diisopropyl phosphoramidite **31**.[109] This final dD cyanoethyl phosphoramidite (dD CE PA) is the form needed to be used on an automated DNA synthesiser in order to incorporate dD into an oligonucleotide strand.

3.2.1.2 Synthesis of dE

To convert the thiomethyl function into the oxopyrimidine, several possible procedures were investigated. As a starting point, a test reaction was performed, following an old literature procedure.[112] In this test reaction, 4-methyl-2-(methylthio)pyrimidine **15** was treated with 32 % HCl solution at 130 °C. The reaction afforded 4-methyl-2-hydroxypyrimidine as its hydrochloride in good yield (*Scheme 3.9*).



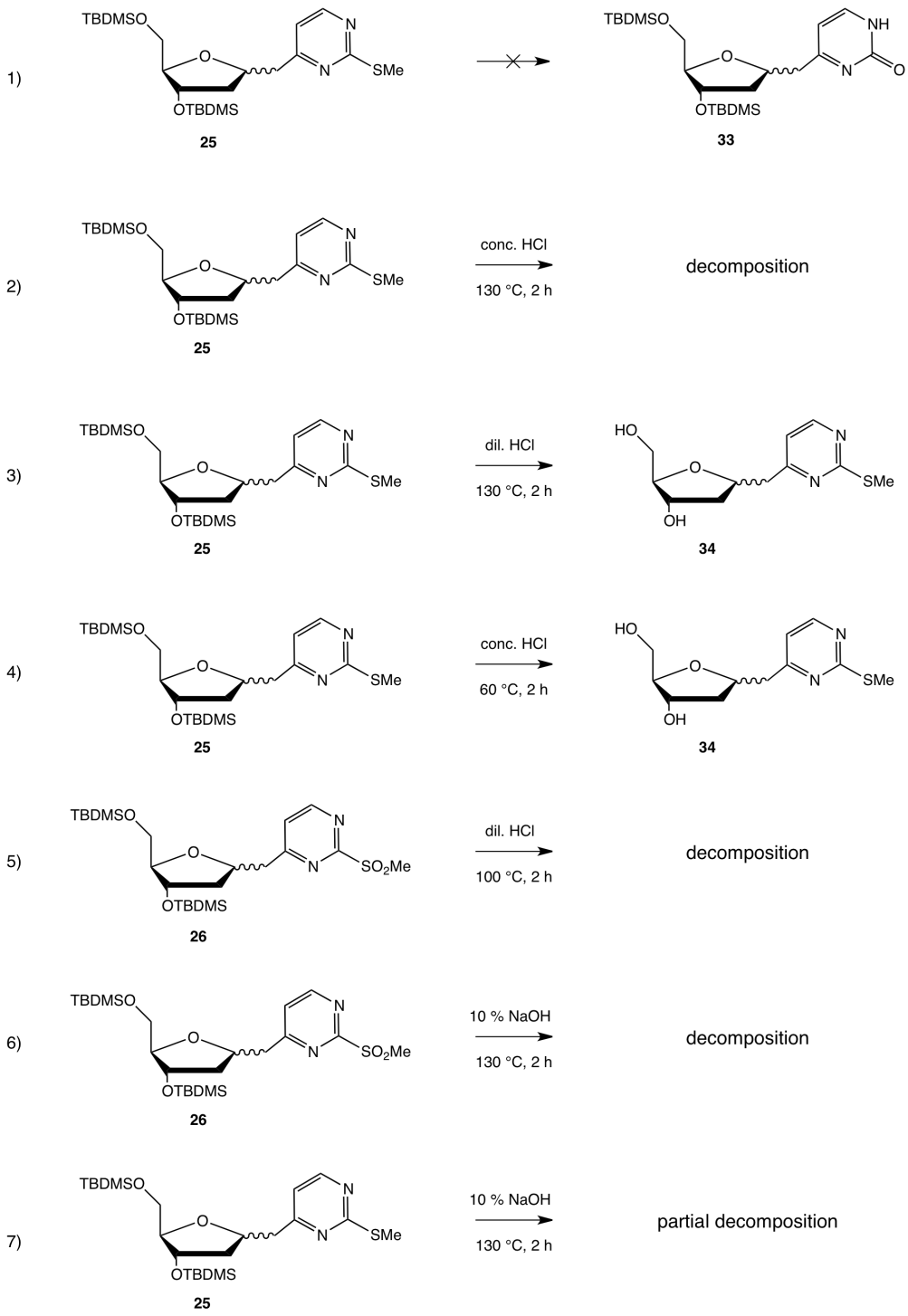
15 **32**

a) HCl, H_2O, 130 °C, 2 h, X%.

***Scheme 3.9**: Test reaction for the functionalisation of the 2-(thiomethyl)pyrimidine to the oxopyrimidine.*

The same reaction conditions were then transferred to compound **25**. Since TBDMS protecting groups are known to be labile to strong acids[113] we expected to obtain deprotected product **33** (*Scheme 3.10*, entry 1). However, the conditions seemed to be too harsh since full decomposition took place (*Scheme 3.10*, entry 2). Using diluted HCl (*Scheme 3.10*, entry 3) or reducing the temperature to 60 °C (*Scheme 3.10*, entry 4), resulted in a deprotection of compound **25**. However, the thiomethyl group did not react. We decided to convert thiomethyl **25** to sulfone **26** first and then try the same reaction conditions again because sulfones are the

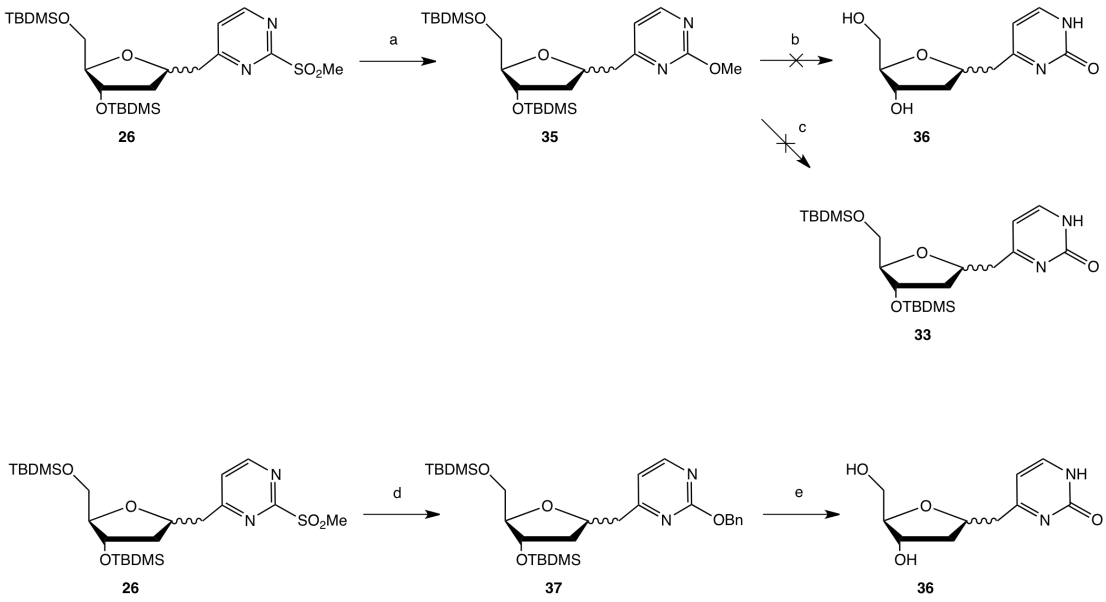
better leaving groups. But this time, only decomposition could be observed (*Scheme 3.10*, entry 5). Additionally, the reaction was tried in the presence of aq. base instead of acid but with no success (*Scheme 3.10*, entry 6). Finally, thiomethyl **25** was treated with aq. base where upon partial decomposition could be observed and some starting material could be re-covered (*Scheme 3.10*, entry 7).



***Scheme 3.10**: Reaction attempts towards dE (33) with different reaction conditions.*

Since all these attempts failed to form dE, we tried a slightly different route with one more step included (*Scheme 3.11*, top). First, sulfone **26** was treated with NaOMe in MeOH to give **35**.[114] The next step should deprotect the methoxy funcion in presence of *lewis* acidic BBr_3 to give the desired ketone[115] but treatment with BBr_3 only led to deprotection of the TBDMS groups and did not react with the methoxy group. Also, a special procedure to deprotect methoxy groups on aromatic rings in presence of TMS-I failed in our hands[116]. The product's mass could be observed by ESI-MS when analysing the reaction mixture, however, the product could never be isolated.

Because methyl groups are generally difficult to cleave when used as protecting groups for alcohols,[117] it was switched to a benzyl protecting group. The synthesis procedure was analogical to the synthesis of methoxy nucleoside **35** but with NaOBn instead of NaOMe as nucleophile. The obtained nucleoside **37** was then treated with BBr_3 according to a literature procedure[115]. The treatment with *Lewis* acid did not only cleave the benzyl protecting group but also the TBDMS groups, so that the free diol was obtained (*Scheme 3.11*). Since the deprotection would be the next step anyway, this saves one step. The rather low yield of 24 % describes the yield at the first try, so there is a high potential for optimisation of the reaction.

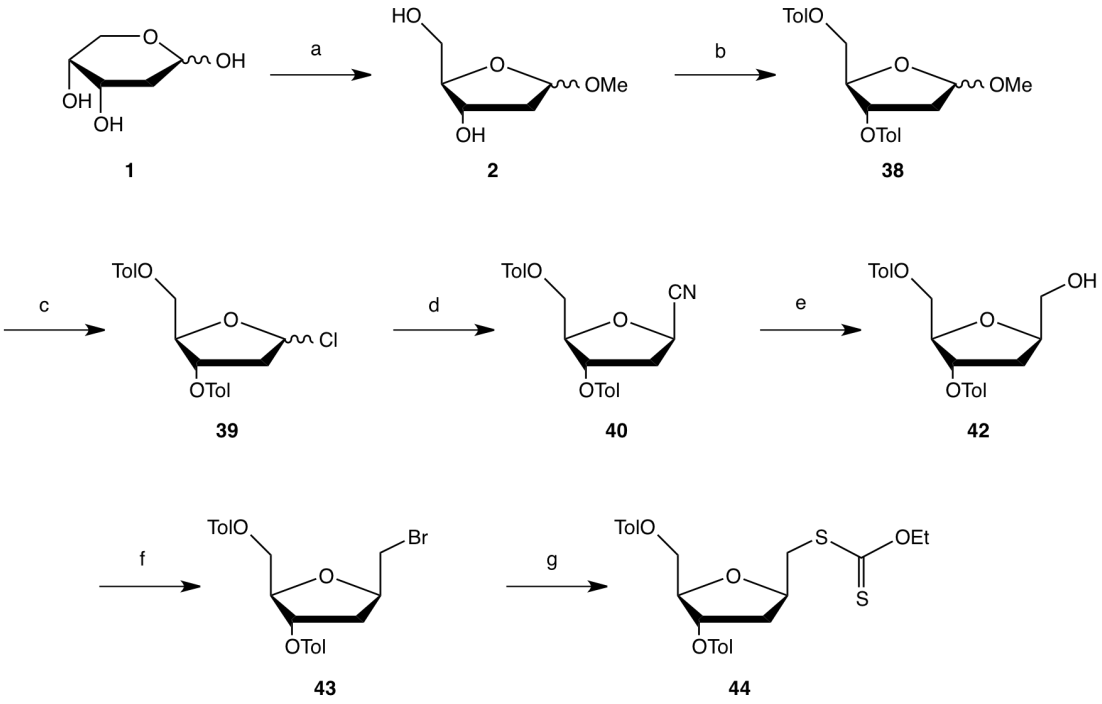


a) NaOMe, MeOH, reflux, 3 h, 78 %. b) BBr_3, CH_2Cl_2, 0 °C, 3 min. c) TMS-I, X. d) NaOBn, BnOH, 70 °C, 2.5 h, 58%. e) BBr_3, CH_2Cl_2, 0 °C, 3 min, 24 %.

***Scheme 3.11**: Unsuccessful (top) and successful (bottom) synthesis of dE (**36**).*

3.2.2 Radical Pathway

As described in *Section 3.1*, a fundamental different pathway to the *Wittig* route was designed as well. In this radical pathway, named after its last synthetic step that includes a radical reaction, the methylene bridge belongs to the sugar moiety and can therefore be fixed in its β-form prior to the coupling to the pyrimidine base. This would circumvent the necessity of time-consuming anomer separation by HPLC.



a) HCl, MeOH, rt, 15 min, 99 %. b) *p*TolCl, pyr, 50 °C, 2 h, 83 %. c) acetyl chloride, AcOH, rt, 5 min, 75 %. d) TMSCN, BF_3*Et_2O, rt, overnight, 53 %. e) i) HCl, dioxane, 60 °C, 6 h; ii) BH_3*DMS, THF, rt, 16 h, 58 %. f) CBr_4, PPh_3, CH_2Cl_2, rt, 1 h, 93 %. g) KSC(S)OEt, acetone, 60 °C, 2 h, 60 %.

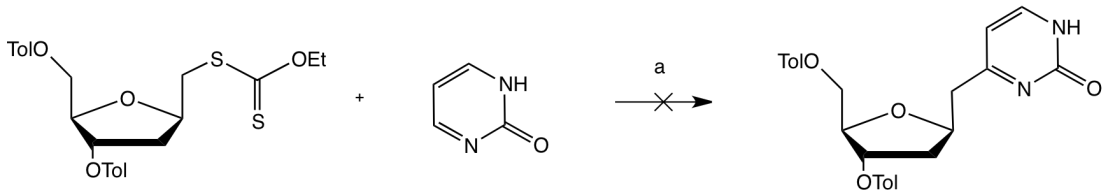
***Scheme 3.12**: Synthesis of the xanthate nucleoside building block.*

Scheme 3.12 shows the synthesis of the xanthate building block, needed for the radical reaction in the final coupling step. It starts from 2-deoxy-D-ribose, which has been converted to acetal **2** almost quantitatively. Protection of the free hydroxy groups with *p*-toluoyl protecting groups gave **38** as an anomeric mixture.[118] For the formation of *Hoffer's* chlorosuger **39**, a literature procedure by *Rolland et al.* was followed.[119] This procedure is based on the original procedure by *Hoffer*[120] but instead of inserting HCl gas it proceeds through in situ generation of HCl which is easier to handle and promises higher yields. The following nucleophilic sub-

stitution to **40** in presence of TMSCN was done according to a procedure by *Grünefeld* and *Richert*.[121] In principle, the reaction could also be performed with NaCN instead of TMSCN but the authors claimed to obtain higher yields with the latter. Nucleoside **40** was obtained in an anomeric mixture of α/β 1:5 and was separable by column chromatography. The synthesis was continued with the pure β-anomer. Conversion to compound **42** was achieved in two steps through an acid intermediate, which was not purified but directly reduced to alcohol **42**.[122] In the following step, **42** was converted to bromide **43** in very good yield. Treatment with potassium ethylxanthate afforded the xanthate sugar building block needed to perform the final radical coupling.[123]

The other building block for the radical reaction, the pyrimidine ring, is commercially available. 2-hydroxypyrimidine hydrochloride and benzoyl protected 2-anminopyrimidine can theoretically be used in order to synthesise dE and dD, respectively.

With the two building blocks in hand, the coupling reaction could be investigated. We followed a procedure by *Osornio et al*.[124], using dilauroyl peroxide (DLP) as radical initiator to perform an oxidative radical alkylation (*Scheme 3.13*). DLP was added slowly by syringe pump over a 12 h period of time. Unfortunately, no coupled product **46** was detected, only starting material could be recovered.



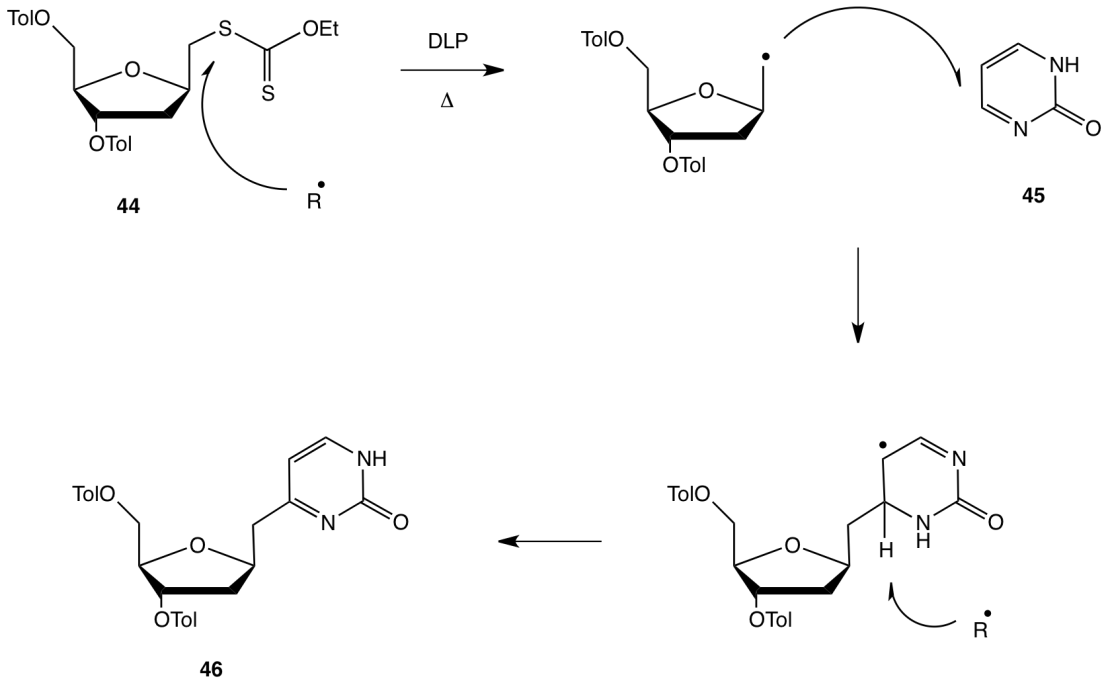
44 **45** **46**

a) DLP, 1,2-dichloroethane, reflux, 12 h.

Scheme 3.13*: Unsuccessful radical coupling reaction.*

In order to proceed, the reaction needs to form a primary radical intermediate, which is known to be very unstable (*Scheme 3.14*). The result of the reaction leads to the conclusion that the primary radical did not form at all and the coupling reaction could therefore not take place. In fact, the xanthates used in literature to perform radical couplings are always secondary xanthates or placed next to a carbonyl function that helps stabilising the radical.[125] Since such a stabilised radical would not lead to the desired product for this project, this synthetic

strategy was not investigated any further. Nevertheless, bromocompound **43** might be a useful reaction partner in another kind of coupling reaction.

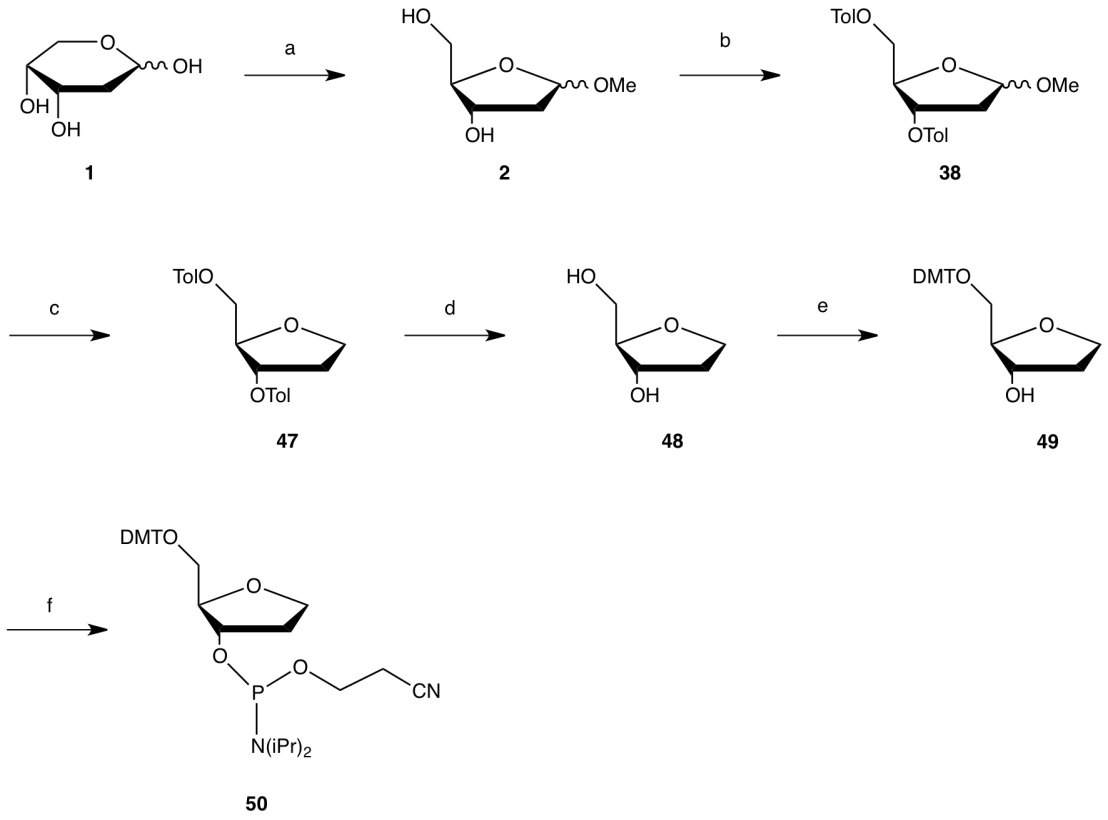


***Scheme 3.14**: Proposed mechanism showing the formation of the unstable primary radical.*

3.2.3 Synthesis of an abasic site nucleoside

An abasic site (As) is a nucleoside that lacks the heterocycle involved in Watson-Crick base pair formation.[126] The cyanoethyl phosphoramidite of a 2-deoxy abasic site was synthesised in order to be incorporated into oligonucleotide strands. It was then used as a reference compound for the binding strength of the nucleoside analogue dD.

Like the syntheses described above, the synthesis of the abasic site starts from 2-deoxy-D-ribose (*Scheme 3.15*). The two first steps to compound **38** are identical with the starting steps of the radical pathway (*see Section 3.2.2*). Acetal **38** was then converted by triethylsilane in the presence of boron trifluoride to **47**, following a literature procedure.[127] It followed the de-esterification to diol **48**[127] and subsequent protection of the 5'-hydroxy group with DMT[111]. Finally, the compound was converted to its cyanoethyl phosphoramidite form (dAs CE PA, **50**).[109]



a) HCl, MeOH, rt, 15 min, 99 %. b) TolCl, pyr, 50 °C, 2 h, 83 %. c) Et_3SiH, BF_3*Et_2O, CH_2Cl_2, 0 °C, 1 h, 86 %. d) NaOMe, MeOH, rt, 6 h, 80 %. e) DMTCl, DMAP pyr, rt, 20 h, 44 %. f) $PCl(N(iPr)_2OEtCN)$, $NEt(iPr)_2$, CH_2Cl_2, rt, 2 h, 83 %.

***Scheme 3.15**: Synthesis of dAs CE PA (**50**).*

3.3 Crystallisation attempts

One of the major drawbacks of the original synthesis strategy (see *Section 2.1*) as well as the *Wittig* pathway (see *Section 3.2.1*) is the fact that the anomeric mixture does not resolve with standard silica column chromatography. The anomers could be isolated but necessitated separation by HPLC, which does not allow for a scale up of the synthesis and even for small scales, it is time-consuming and costly. Another difficulty of the work with the homo-*C*-nucleosides in this project is the fact that they have in most cases the appearance of an oil. The reason for this oily consistence is probably the methylene bridge, which induces an additional degree of freedom to the molecule and makes it more flexible and therefore less accessible to crystallisation. Oils are generally more difficult to handle, when it comes to drying of the compound, for instance.

It would be very instrumental to find a way to crystallise one or both anomers. First of all, a crystal structure would allow to gain insight into the three dimensional structure of the molecules. Then, with one of the anomers being more likely to precipitate from a solution, separation would be facilitated. Even if such a fractional crystallisation is not possible, the crystalline form might change the elution properties of the anomers on a silica column in a way that separation will become possible. Finally, solids are easier to handle than oils.

For all these reasons, different strategies were investigated with the overall goal of obtaining crystals of homo-*C*-nucleoside dD or its derivatives. The general approach was to add different protecting groups to the molecule, some of them known to have good crystallisation properties, to test for the consistency and to make the molecule bigger in order to facilitate the separation on a silica column. There are two functional groups on compound dD where protecting groups can be added, the hydroxy functions on the sugar moiety or the amino group on the pyrimidine ring. We started with screening different groups on the pyrimidine ring, either protecting groups for the amino function or substitution of the amino group (*Figure 3.3*).[113] While the first four compounds were synthesised in this work, the Fmoc protected compound was described by *Steinauer* in her master thesis.[103]

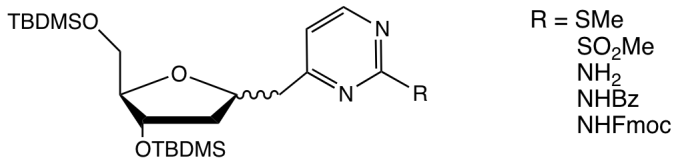


***Figure 3.3**: Variation on the pyrimidine R-group for crystallisation purposes.*

For all the different R-groups, the compound was always an oil and no improvement of anomer separation on a silica column could be observed. Even with the spacious Fmoc protecting group, no change in elution was detected.

Therefore, it was decided to focus on protecting groups on the hydroxy function of the sugar moiety. The bridging silicon ether, shown in *Figure 3.4*, was described before by *Steinauer*[103] and has also an oily texture.

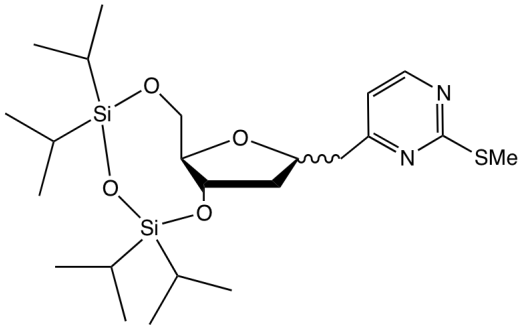


***Figure 3.4**: Compound with bridging silicon ether protecting groups, linking the 3'- and 5'-hydroxy function.*

Since silicon containing protecting groups are known to have poor crystallisation properties[113], focus was placed on other, non-silicon containing protecting groups with ether or ester functions.

Delaney and *Greenberg*[99] reported on separation of their homo-*C*-nucleosides after the protection step of the primary 5'-hydroxy function by dimethoxytrityl (DMT), a very spacious protecting group. We synthesised two 5'-DMT protected compounds (*Figure 3.5*), one with a thiomethyl group on the pyrimidine ring and the other with a benzoyl protected amino function, but anomer resolution was not obtained for either compound. Also, both were oils and we did not succeed in crystallising them.

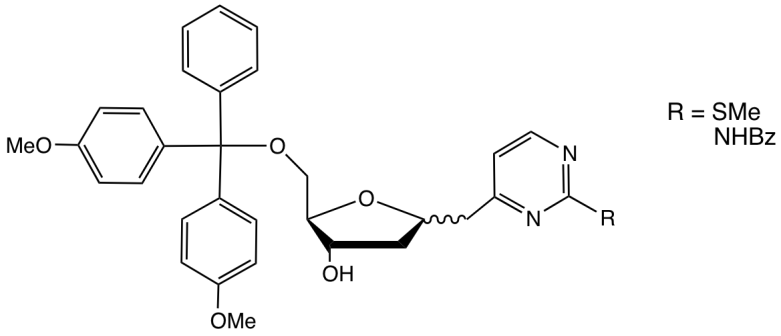


***Figure 3.5**: 5'-DMT protected compounds.*

Finally, we screened four different protecting groups, shown in *Figure 3.6*. The first two, 4-bromobenzoyl **52** and 4-nitrobenzoyl **53**, are known to facilitate crystallisation. 2-Naphtoyl **54** was chosen due to its increased π-surface and 3,5-dimethylbenzoyl **55** is a bulky

protecting group without many conformational degrees of freedom. They were all synthesised from thiomethyl diol **51** in one protecting step by treatment with the corresponding acid chloride in pyridine on the basis of literature procedures.[127,128]

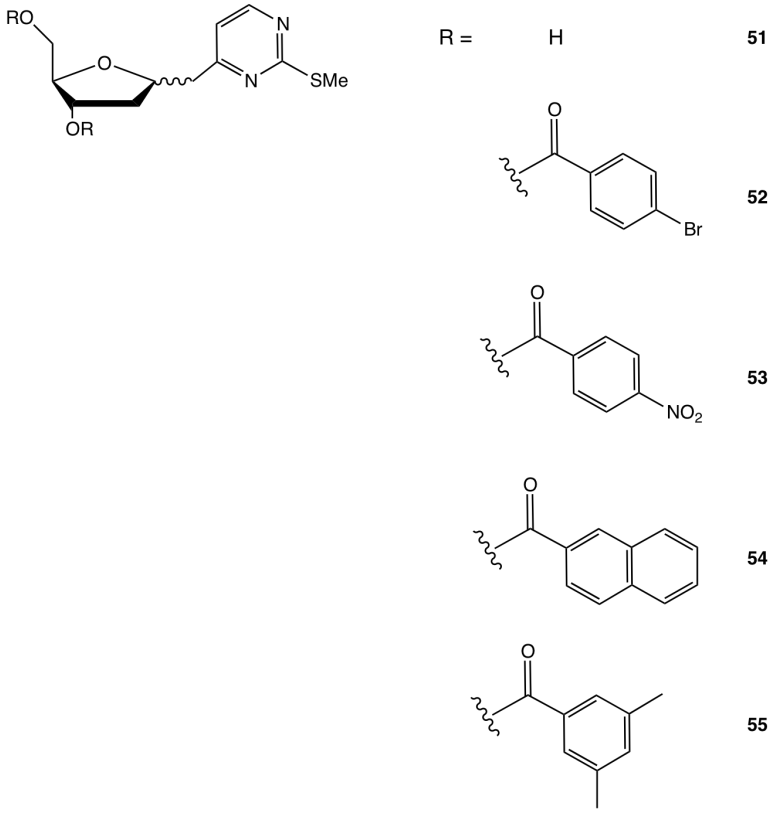
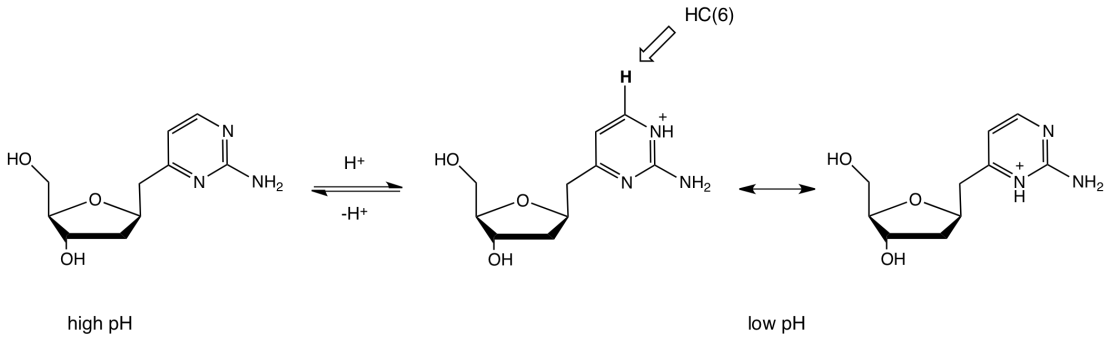


***Figure 3.6**: Screening of different protecting groups for crystallisation purposes: 4-bromobenzoyl, 4-nitropenzoyl, 2-naphtoyl and 3,5-dimethylbenzoyl.*

The nitrobenzoyl protected compound **53** was obtained as an oil and did not show any anomer separation on column. Therefore, it was focused on the three other compounds which where obtained in a solid or wax-like consistency. A solubility screening was performed in order to find possible crystallisation solvents. Finally, crystallisations were set up from the following solvent mixtures: CH_2Cl_2-hexane, CH_2Cl_2-MeOH, CH_2Cl_2-H_2O, MeCN-hexane, MeCN-H_2O, Et_2O-hexane, Et_2O-MeOH and Et_2O-H_2O. Unfortunately, despite all these attempts, no crystals have been obtained, so far.

3.4 pK_a Determination

The acidity constant (pK_a) of dD (**9**) was determined by a pH-dependent ^1H NMR measurement in D_2O. The solution with a 2.5 mM concentration of dD in D_2O was first adjusted to a low pD by the addition of dil. DCl in D_2O, then small aliquots of dil. NaOD in D_2O were added to increase the pD. After the addition of every aliquot, the pD was measured and a ^1H-NMR spectrum was recorded. The pD of the D_2O solution was obtained by adding 0.40 to the pH meter reading. As expected, protonation led to a downfield shift for H-signals close to the protonation site (*Scheme 3.16*). In *Table 3.1*, the results of the titration are summarised, focusing on the HC(6) signal next to the protonation site.



***Scheme 3.16**: dD (**9**) with marked position HC(6) in its unprotonated (left, high pH) and protonated (right, low pH) state.*

pH	pD	H-6	pH	pD	H-6
1.39	1.79	8.294	7.15	7.55	8.213
1.46	1.86	8.295	7.48	7.88	8.213
1.60	2.00	8.296	8.45	8.85	8.213
2.06	2.46	8.296	8.99	9.39	8.212
2.73	3.13	8.291	9.54	9.94	8.212
3.46	3.86	8.280	10.03	10.43	8.211
4.24	4.64	8.255	11.05	11.45	8.213
4.40	4.80	8.247	11.84	12.24	8.211
4.93	5.33	8.226	12.45	12.85	8.211
5.54	5.94	8.212	12.94	13.34	8.209
6.38	6.78	8.214			

***Table 3.1**: pH, pD and its corresponding H-6 shift in ppm.*

The changes in the chemical shifts were evaluated by a Newton-Gauss nonlinear least-squares curve-fitting procedure. The relationship between the varying pD values and the observed chemical shift is describe by *Equation 3.1*.[129]

$$\delta_{obs} = \frac{\delta_{dD} + \delta_{dDH} \cdot 10^{pK_{a/D_2O} - pD}}{1 + 10^{pK_{a/D_2O} - pD}} \tag{3.1}$$

In *Equation 3.1*, δ_{dD} and δ_{dDH} represent the chemical shift of the compound dD in its deprotonated and protonated form, respectively. $pK_{a/D2O}$ corresponds to the negative logarithm of the acidity constant in D_2O. The $pK_{a/D2O}$ value was then converted to the $pK_{a/H2O}$ value, describing the situation in H_2O, by using *Equation 3.2*.[130]

$$pK_{a/D_2O} = 1.015 \cdot pK_{a/H_2O} + 0.45 \tag{3.2}$$

In *Figure 3.7*, the variation of the chemical shift for the C(6) proton in dependence on the pD is shown.

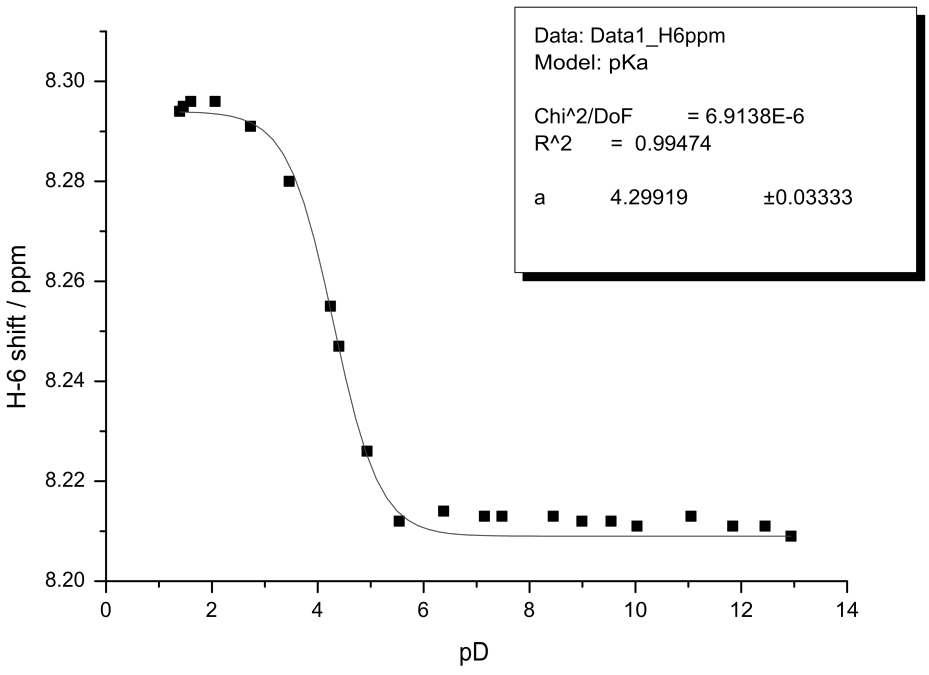


***Figure 3.7**: pD dependence of the ^1H chemical shifts of HC(6) of dD.*

From the curve fitting, the pK_a value obtained is 4.299±0.033 in D_2O. Using *Equation 3.2*, the pK_a in H_2O could be determined as 3.792. This acidity constant can now be compared to the pK_a values of the natural occurring deoxynucleosides, shown in *Figure 3.8* (see also *Table 1.2* for comparison).

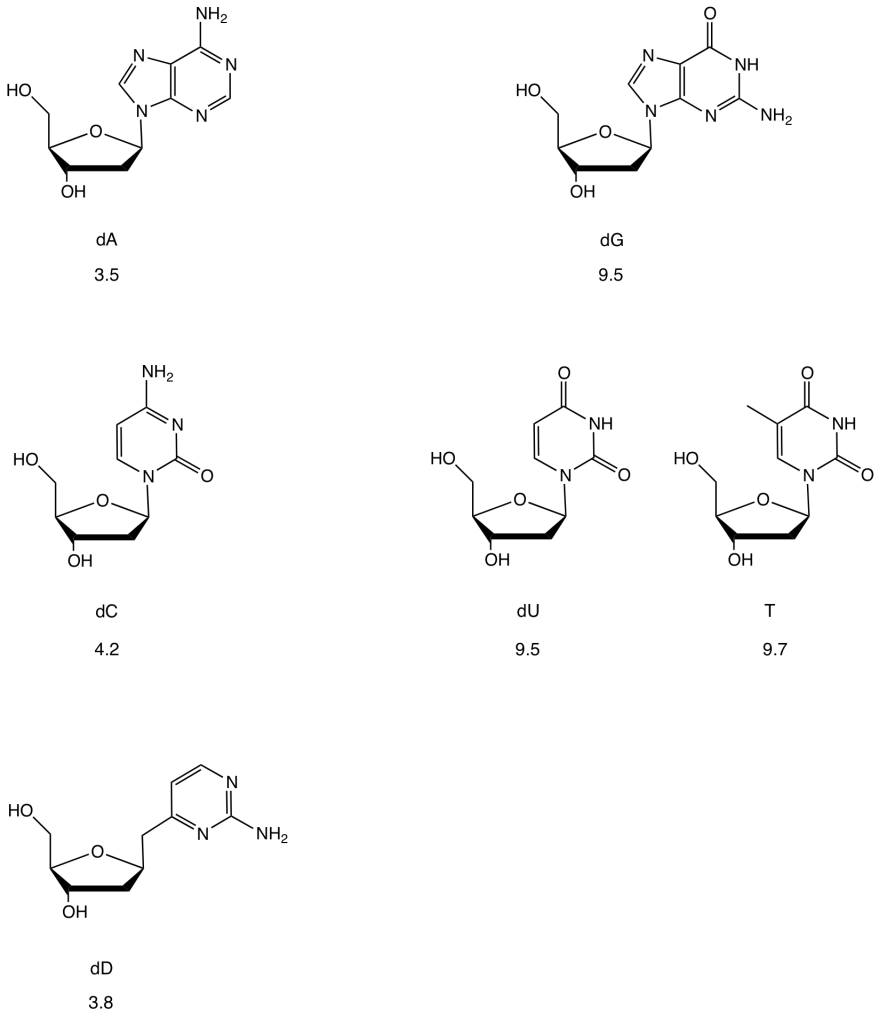


***Figure 3.8**: Natural nucleosides and dD with its pK_a values.*

The pK_a values of the natural deoxynucleosides are taken from the literature.[58] According to their acidity constants, two groups of natural deoxynucloesides can be formed: those with higher pK_a values around 9, including dG, dU and T, and those with lower pK_a values around 4, including dA, dC and dD. The nucleoside with a pK_a value closest to the pK_a of dD is dA with a value of 3.5. This finding supports our hypothesis that dD is an analogue of dA.

3.5 ITC measurements

3.5.1 Oligonucleotides

To test our hypothesis that dD mimics dA, we incorporated it into oligonucleotides in order to perform base pairing studies. The non-self-complementary 13-mer oligonucleotide system (*Figure 3.9*) was described before by *Breslauer* in a different context.[131]

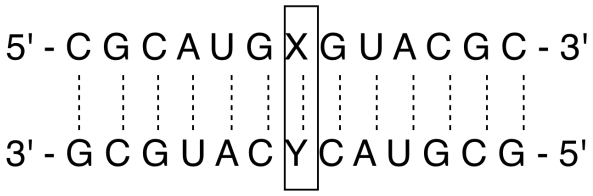


***Figure 3.9**: Investigated complementary oligonucleotides with varying middle positions.*

The positions X and Y can be modified. At these middle positions, the natural deoxynucleosides dA, dG, dC and dU as well as dD and an abasic site should be incorporated, leaving 12 different strands to be synthesised, six of each kind. The abbreviations for the different oligonucleotides are listed in *Table 3.2*.

Strand	Oligonucleotide	Abbreviation
Strand "C"	5'-CGCAUG**U**GUACGC-3'	C_U
	5'-CGCAUG**C**GUACGC-3'	C_C
	5'-CGCAUG**A**GUACGC-3'	C_A
	5'-CGCAUG**G**GUACGC-3'	C_G
	5'-CGCAUG**D**GUACGC-3'	C_D
	5'-CGCAUG**As**GUACGC-3'	C_As
Strand "G"	5'-GCGUAC**U**CAUGCG-3'	G_U
	5'-GCGUAC**C**CAUGCG-3'	G_C
	5'-GCGUAC**A**CAUGCG-3'	G_A
	5'-GCGUAC**G**CAUGCG-3'	G_G
	5'-GCGUAC**D**CAUGCG-3'	G_D
	5'-GCGUAC**As**CAUGCG-3'	G_As

***Table 3.2**: Abbreviations for the synthesised oligonucleotide strands.*

Of all different combinations of complementary strands, ITC measurements were performed, leading to 36 different measurements. The results of these measurements give us a

complete set of thermodynamic data and enable us to draw conclusions of the base-pairing abilities of homo-*C*-nucleoside dD in comparison to all other bases.

Oligonucleotide synthesis was achieved on an automated DNA synthesiser using the *Phosphoramidite* approach, explained in *Section 1.4*. For more details on the synthesis procedure see *Section 5.2.40*.

3.5.2 Student's t-test

To test the statistical significance of the data, student's t-tests were performed. Generally, a t-test is applied for normal populations in cases where samples are small.[132] First, a t-test was made with all the measurements in which one strand contains an abasic site in the middle position against all the other measurements. Y_1,..., Y_9 represent the measurements with one strand containing an abasic site in the middle position and X_1,..., X_{38} represent the measurements where no abasic site was included. μ_x and μ_y are the means of the population from which these samples are drawn. The null and alternate hypotheses are:

$$H_0: \mu_x - \mu_y \leq 0 \qquad \text{vs.} \qquad H_1: \mu_x - \mu_y > 0$$

with the alternate hypothesis saying that the results of the measurements of abasic site containing strands are significantly lower then the results of all the other measurements and the null hypothesis saying that the results of abasic site containing measurements are not significantly lower.

Using the averages and standard deviations from *Table 3.3*, a t-value of 1.856 is obtained. Consulting the t table, a P-value between 0.05 and 0.025 is found. If we follow the 5% rule, H_0 can be rejected and μ_y (containing abasic site) is significantly lower than μ_x.

	Average	Stdev
All	-210	67
U	-247	67
C	-214	85
A	-245	74
G	-214	54
D	-176	48
As	-163	48

***Table 3.3**: Summery of all means and standard deviations (in kJ/mol).*

A second t-test was made with Y_1,..., Y_{12} representing the measurements with one strand containing the artificial nucleotide D in the middle position and X_1,..., X_{38} representing all measurements with natural nucleotides in the middle position.

The null and alternate hypotheses are

$$H_0: \mu_x - \mu_y \leq 0 \qquad \text{vs.} \qquad H_1: \mu_x - \mu_y > 0$$

The obtained t-value is 0.682, corresponding to a P-value between 0.25 and 0.40, meaning that we cannot reject the null hypothesis. That does not mean that we can assume the H_0 to be true but we have evidence that the different means of the two populations is due only to random variation.

In summary, we can assume a significant difference between nucleosides with a nucleobase and those without. Furthermore, there seems not to be significant difference between the strands incorporating D in the middle position compared to those with another nucleobase in the middle position. Hence, we can assume that nucleoside D shows certain binding properties comparable to those of the natural nucleosides.

3.5.3 Data analysis

The process of duplex formation can roughly be divided into two parts. First, the DNA single strands form a pre-ordered structure by stacking. Second, the actual duplex formation takes place by binding of the two pre-stacked single strands. As pointed out by *Vesnaver* and *Breslauer*, the pre-ordered single strand already posses over 40 % of the total enthalpy.[131] The contribution from single-stranded structures is included in the enthalpies measured by ITC.

Most of the base pair combinations were measured once. Nevertheless, some measurements were repeated to check the precision of the measurements. For example, the combination with A in the middle positions of both complementary strands was measured twice. In the first measurement, the C-strand was titrated to the G-strand in the cell of the calorimeter. This gave an enthalpy of -243 kJ/mol. In the repeating measurement, the strands were swapped and the G-strand was now titrated to the C-strand, yielding an enthalpy of -241 kJ/mol, which is a difference of less then 1%. The same was done with the combination containing D in both middle positions. In the first measurement, the C-strand was titrated to the G-strand and an enthalpy of -155 kJ/mol was obtained. In the second measurement, the G-strand was added to

the C-strand, which gave an enthalpy of -156 kJ/mol. Again, this is a difference of less than 1 %.

In *Table 3.4*, the obtained enthalpies in kJ/mol of all ITC measurements are summarised. All measurements can be found in the appendix.

		G					
		U	C	A	G	D	As
C	U	-294	-285	-397	-227	-221	-176
	C	-177	-152	-178	-181	-153	-111
	A	-274	-384	-243	-204	-210	-167
	G	-253	-341	-243	-217	-211	-177
	D	-201	-188	-216	-167	-155	-114
	As	-161	-271	-188	-127	-123	

***Table 3.4**: ΔH values in kJ/mol.*

To check the accuracy of the measurements, the data was compared to a very similar system measured by *Breslauer* et al. They worked with the same sequences which were used in this project with the only difference, that their sequences contained T instead of dU. For the A:U base pair with the A in the C-strand and the U in the G-strand, we measured ΔG of -47.9 kJ/mol, ΔH of -274 kJ/mol and ΔS of -758 J/mol*K. *Breslauer* et al. obtained their thermodynamic profiles not by ITC but by differential scanning calorimetry (DSC) and temperature-dependent UV absorbance spectroscopy. They observed ΔG of 83.7 kJ/mol, ΔH of 490 kJ/mol and ΔS of 1362 J/mol*K.[131] These values are in the same range like the values we observed and therefore comparable. The differences in the values probably base on the different techniques used to determine the thermodynamic profile.

The *Watson-Crick* base pairs are expected to show the highest binding energies, thus the most negative enthalpy values. As can be seen in Table 3.4, they are indeed among the strongest, except for the G-C *Watson-Crick* base pair with C in the middle position of the C-strand opposite G in the middle position of the G-strand. Rather high binding energies were also found for the G:U base pair, which forms a *Wobble* base pair similar in energy to the *Watson-Crick* base pairs (described in *Section 1.2.4.3*).

dD seems to show a certain selectivity for dU. Focusing on the G_D strand, one can see that dD shows highest binding energies with an enthalpy of -221 kJ/mol for dU in the C-

strand. The second highest energies are observed when dD is opposite dG and dA with enthalpies of -211 and -210 kJ/mol, respectively. On third place is the binding towards dD and dC with -155 and -153 kJ/mol, respectively. As expected, binding towards the abasic site exhibits the lowest binding energy. When focusing on the C-strand with dD in the middle position, only the binding towards dA with -216 kJ/mol is stronger than towards dU with -201 kJ/mol. It follows in order of decreasing binding enthalpies: dC, dG, dD and finally dAs. Binding of dD towards dG and dA in comparison of the binding pattern of dD with dU is shown in *Figure 3.10*. Binding towards dA can be explained by rotation of the nucloebase around the methylene bridge. In summary, a binding affinity of dD for dU can be observed. Additionally, dD shows also a certain binding affinity for dA and dG. The same observations were also found by *Bischof* when he analysed the same systems by melting curves.[98] He found a slight selectivity of dD for dU but also for dA and, in one case, for dG.

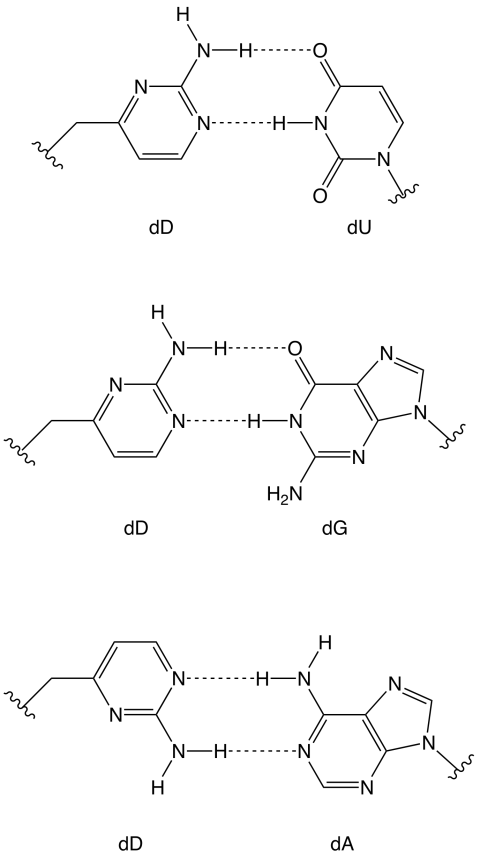


***Figure 3.10**: Possible binding patterns of dD to natural deoxynucleosides.*

dD is presumably an analogue of dA, therefore they should exhibit similar base pairing behaviour. *Table 3.5* summarises the comparison of the enthalpies of dD and dA in kJ/mol.

On the left side, the data for dD and dA in the G-strand opposite all different nucleosides in the C-strand are shown. The right side table shows dD and dA in the C-strand opposite all different nucleosides in the G-strand. The values are in decreasing binding strength order for dD.

		G		
		D	A	Δ
C	U	-221	-397	176
	G	-211	-243	32
	A	-210	-243	33
	D	-155	-216	61
	C	-153	-178	25
	As	-123	-188	65

		C		
		D	A	Δ
G	A	-216	-243	27
	U	-201	-274	73
	C	-188	-384	196
	G	-167	-204	37
	D	-155	-210	55
	As	-114	-167	53

Table 3.5*: Comparison of ΔH values of dD and dA in (kJ/mol).*

Compared to the series with dA in the middle position, dD shows generally lower binding energies. This can be explained by the additional degree of freedom that the methylene bridge of homo-*C*-nucleosides introduces. The extra flexibility can decrease the binding energy as well as disrupting the base stacking. Nevertheless, the trends observed in binding of dA are similar to the ones of dD. The binding strengths of A in the G-strand are in decreasing order: dU, dG and dA, dD, dC. Exactly the same order can be observed for G_D. Not so distinct but similar is the binding order for C_A compared with C_D. Thus, we can conclude that a certain similarity between the binding behaviour of dD and dA can be observed.

The difference in enthalpy between dD and dA is smaller when they are in the middle position of the G-strand. When they are in the C-strand, a larger difference is observed. In other words, the binding behaviour of dD is closer to the one of dA when they are in the G-strand, flanked on both sides of a cytosine. This effect was already observed by melting curve analysis of the same oligonucleotides and could be explained by the fact, that cytosine is known to induce strong dipoles to its nearest neighbours which fixes the flexibility of the homo-*C* nucleoside dD[98]. This stacked conformation is then similar to the one of the non-flexible dA.

The entropies and Gibbs energies of the systems cannot be measured directly by ITC. Nevertheless, they can be calculated by *Equations 1.1* and *1.2* as described in *Chapter 1.5*.

The ΔS values and ΔG values are summarised in *Table 3.6* and *Table 3.7*, respectively. ΔS values are shown in J/mol*K while ΔG values are in kJ/mol.

		G					
		U	C	A	G	D	As
C	U	-838	-811	-1165	-620	-607	-452
	C	-455	-367	-454	-452	-380	-232
	A	-758	-1139	-667	-540	-570	-422
	G	-704	-978	-657	-586	-578	-465
	D	-539	-498	-590	-424	-392	-248
	As	-404	-773	-494	-286	-284	

***Table 3.6**: ΔS values in J/mol*K.*

		G					
		U	C	A	G	D	As
C	U	-44.1	-43.2	-49.7	-42.3	-40.0	-41.1
	C	-41.4	-42.6	-42.5	-46.3	-39.8	-41.7
	A	-47.9	-44.3	-44.1	-42.9	-40.1	-41.1
	G	-43.2	-49.4	-47.1	-42.2	-38.8	-38.5
	D	-40.2	-39.5	-40.0	-40.6	-38.2	-39.9
	As	-40.5	-40.5	-40.6	-41.6	-38.3	

***Table 3.7**: ΔG values in kJ/mol.*

The obtained free energies are in the range of -38.2 to -49.7 kJ/mol. These values are small compared to the measured enthalpies of -111 to -397 kJ/mol and the entropic term (-TΔS) of 69.1 to 347 kJ/mol. While the duplex formation is highly favoured by enthalpy, it is nearly as much unfavoured by entropy. These two opposing contributions are much larger than the resulting free energy which stabilises the duplex only slightly. This effect is known in literature for other oligonucloetides[133,134].

Like in the case of the enthalpies, *Watson-Crick* base pairs are among the most favourable in terms of free energy. The five most negative ΔG values come from the four *Watson-Crick* base pairs as well as the G:A combination in the middle position. The later might result from the three neighbouring deoxyguanosines in the middle position of the C-strand, that exhibit high stacking energies.

Even if the ΔG values are lower and the resulting trends therefore less pronounced then for the ΔH values, the conclusions for dD are similar like those drawn from the enthalpies. It must be noted that all of the ΔG values are in the same range and rather close together. For the series with dD in the middle position of the G-strand, the most favourable free energies are those of the measurement opposing A with -40.1 kJ/mol and the measurement opposing U with -40.0 kJ/mol. Then, it follows in decreasing order C, G, As and D. The order for dD in the C-strand is slightly different. The most favourable free energy occurs for G in the opposite position, with a ΔG value of 40.6 kJ/mol, followed by U with 40.2 kJ/mol. Then, in decreasing order, it follows A, As, C and D. Like in the comparison of the enthalpies, we can observe a slight selectivity of dD towards, dU, dA and dG in terms of free energy.

4 Outlook

As shown in the previous chapter, dD exhibits a certain ability to mimic dA. This finding opens many new possibilities. A particularly interesting field would be PCR-like experiments. Two different sets of experiments are possible. In one set, a template strand with dD incorporated would be mixed with a primer, a polymerase and the natural triphosphates ATP, GTP, CTP and UTP. Following the hypothesis that dD mimics dA, the polymerase would incorporate dU opposite dD. In the other set of experiments, DTP could be mixed with different template strands, a primer and a polymerase to observe from the opposite of which nucleoside would the polymerase incorporate dD. According to the hypothesis, it should be incorporated opposite dU, but not opposite of the other nucleosides.

In this work, we managed to synthesis homo-*C*-nucleoside dE for the first time (see *Section 3.2.1.2*). With this synthetic strategy, it is now possible to synthesise the phosphoramidite form of dE, to incorporate it into oligonucleotide strands and, as done in this work with dD, to determine its thermodynamics by ITC. It would allow for evaluation of an all-pyrimidine system.

A crucial point of the presented hypothesis is the synthesis of dD as well as dE under prebiotic conditions, which was not part this project. Nevertheless, it should be investigated in order to establish the EANs dD and dE as precursor of present-day nucleosides.

In principle, every biochemical process that uses adenosine could be a possible target for exchange of dA with dD. Examples are ATP, CoA, NAD^+ or $NADP^+$. Additionally, it might be worthwhile to test for antibiotic or antiviral activities of dD, since many EANs like Clitocine and also some *C*-nucleosides show such activities and are used as active ingredient in drugs (see *Section 1.3.1*).

5 Experimental Part

5.1 General

Reagents and Solvents: All reagents were used as purchased from commercial suppliers unless otherwise stated. For reactions, solvents of pro analysis grade were used. For reactions performed under dry atmosphere, solvents of puriss. grade were used. For work up and purification, distilled solvents of technical grade were used.

Chromatography: Thin layer chromatography (TLC): *Merck* TLC aluminium sheets, silica gel *60* F_{254}, 2 mm. Column chromatography (CC): *Sigma-Aldrich* Silica gel Merck Type 9385, 230-400 mesh, 60 Å.

High-performance liquid chromatography (HPLC): Normal phase: Analytical HPLC: *Shimadzu LC-10AT,* Spherisorb®-NH_2, 5 µm, 25 cm x 4.6 mm. Reverse-phase (RP): YMC basic, 5 µm, 15 cm x 10 mm. Preparative HPLC: *Shimadzu LC-8A*, Spherisorb®-NH_2, 5 µm, 25 cm x 20 mm.

Melting Points (M.p.): Heating microscope from Christoffel Labor- und Betriebstechnik; Melting points are uncorrected.

Infrared Spectroscopy (IR): *Jasco* 4100 FT-IR spectrometer; absorption values in cm^{-1} and intensity (*s: strong*, 0-30 % transmission; *m: middle*, 30-60 % transmission; *w: weak*, 60-100 % transmission).

Nuclear Magnetic Resonance (NMR): ^{1}H-, ^{13}C and ^{31}P-NMR: *Bruker AV-300, AV-400* or *AV-500* instruments. ^{13}C-signal multiplicity was deduced from DEPT 90 and DEPT 135 spectra (*Disortionless Enhancement by Polarization*). Peak assignment was performed by two-dimensional NMR experiment (NOESY, COSY).

Mass Spectrometry (MS): *Finnigan Trace GC ultra* instrument equipped with a Zebron ZB-5MS capillary GC column for CI and EI; *Finnigan Surveyor MSQ* quadrupole spectrometer for ESI; *m/z* (rel. %); *Bruker* Autoflex I spectrometer for MALDI-TOF (matrix-assisted laser desorption/ionisation time of flight).

pK_a Titration: Acidity constants were determined by pH-dependent ^{1}H NMR measurement on a *Bruker* AV-500 insturment in D_2O. The change of the chemical shift of HC(6)

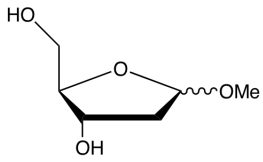
was evaluated in dependence on adjustement of the pD values by DCl and NaOD. The experiment was performed at 2.5 mM concentration in 0.7 ml D_2O (298 K, I = 0.1 M $NaNO_3$). The centre peak of tetramethylammonium ion (0.5 mM; δ = 3.1776 ppm rel. to TMS) was used as internal reference. The pH was measured with a Hamilton Minitrode glass electrode connected to a Metrohm 605 digital pH meter. 0.4 log units were added to the pH-meter reading to obtain the corresponding pD of the D_2O solution. The change in chemical shift of the HC(6) proton in dependence on the pD was evaluated by a Newton-Gauss non-linear least-squares curve-fitting procedure.

Isothermal Titration Calorimetry (ITC): All batches were dialyzed previous to any use against a large volume of milli-Q H_2O with a MW cutoff of 1 kDa. The oligonucleotide concentrations were determined spectrophotometrically by measuring the absorbance. A pH = 7 10 mM sodium phosphate buffer with 1.0 M NaCl and 1 mM EDTA, was used. ITC experiments were performed at 25 °C on a VP-ITC MicroCalorimeter from MicroCal. All samples were degassed under vacuum prior to the titration experiments. Aliquots of 10 µl of a solution of one of the strands of a known precise concentration (in the 50 µM range) were injected into the 1.4 ml titration cell containing the solution of the complementary strand (in the 5 µM range). 25 injections were typically performed for each set of experimental conditions at intervals of 240 s to ensure the return to equilibrium. Blank experiments were performed in order to take ligand dilution heat into consideration (heat removed before analysing the titration data). The dilution heats were identical to the heat signals detected after saturation was reached. Data analysis were performed using a nonlinear least-squares fitting algorithm software (MicroCal Origin 7.1) with a 1:1 binding. The heat of the first injection was systematically removed for analysis.

UV Spectra: Agilent 8453UV-Vis spectrophotometer.

5.2 Synthesis

5.2.1 Methyl-2-deoxy-α/β-D-ribofuranosid (2)

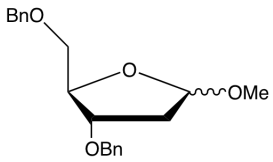


2-Deoxy-D-ribose **1** (13.65 g, 101.77 mmol) was solved in 250 ml MeOH. 28 ml of 1% HCl in MeOH was added. After stirring for 50 min at rt, silver carbonate (4.98 g, 18.10 mmol) was added and it was filtered through celite. The solvent was evaporated under reduced pressure and it was dried at the HV overnight at 40 °C. A yellow oil (14.47 g, 99 %) was obtained.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 5.14 (*d*, *J* = 2.1 Hz, 1 H, HC(1), β-**2**), 5.11 (*d*, *J* = 4.0 Hz, 1 H, HC(1), α-**2**), 4.16-4.11 (*m*, 1 H, HC(3), α/β-**2**), 4.08-4.04 (*m*, 1 H, HC(4), α/β-**2**), 3.74-3.58 (*m*, 2 H, $H_2C(5)$, α/β-**2**), 3.39 (*s*, 3 H, H_3CO, α/β-**2**), 2.18-1.89 (*m*, 2 H, $H_2C(2)$, α/β-**2**).

^{13}C-NMR (75 MHz, $CDCl_3$): δ 105.3, 105.2 (C(1), α/β-**2**), 87.0, 86.7 (C(4), α/β-**2**), 72.3, 71.7 (C(3), α/β-**2**), 64.7, 63.4 (C(5), α/β-**2**), 55.2, 54.8 (OCH_3, α/β-**2**), 41.9, 41.2 (C(2), α/β-**2**).

5.2.2 Methyl-3,5-di-*O*-benzyl-2-deoxy-α/β-D-ribofuranosid (3)



To a solution of **2** (14.47 g, 97.66 mmol) in 200 ml dry THF under Ar, NaH (8.78 g, 365.83 mmol) was added under ice cooling, followed by the addition of $IN(Bu)_4$ (0.33 g, 0.91 mmol) and benzylbromide (22 ml, 31.68 g, 185.21 mmol) and it was stirred overnight at rt.

It was filtered and Et_2O was added. The ether phase was washed with sat. $NaHCO_3$ solution, sat. NaCl solution and water. It was dried over Na_2SO_4, the solvents were evaporated and it was purified by column chromatography (SiO_2, Hex/Et_2O 3:1) to give **3** (19.02 g, 58 %) as a yellow oil.

R_f (SiO_2, Hex/Et_2O 3:1): 0.24, 0.35 (α- and β-anomer)

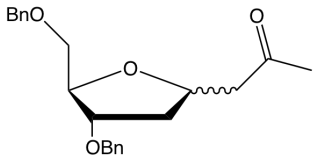
IR (film): 3648*w*, 2081*m*, 3060*s*, 3026*s*, 3001*m*, 2922*s*, 2851*s*, 2336*w*, 1943*w*, 1871*w*, 1804*w*, 1747*w*, 1667*w*, 1601*m*, 1583*w*, 1540*w*, 1493*s*, 1452*s*, 1365*m*, 1328*w*, 1313*w*, 1207*w*, 1180*m*, 1154*m*, 1096*s*, 1049*s*, 1028*s*, 966*m*, 907*m*, 841*w*, 754*s*, 702*s*, 619*w*, 540*m*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.33-7.23 (*m*, 10 H, arom. H, α/β-**3**), 5.06 (*ddd*, *J* = 7.0, 5.3, 1.7 Hz, 1 H, HC(1), α/β-**3**), 4.55-4.44 (*m*, 4 H, benz. H, α/β-**3**), 4.24 (*tt*, *J* = 7.1, 3.6 Hz, 1 H, HC(4), α/β-**3**), 4.12 (*ddd*, *J* = 6.7, 5.9, 3.8 Hz, 1 H, HC(3'), α-**3**), 3.95 (*ddd*, *J* = 7.9, 4.4, 2.8 Hz, 1 H, HC(3'), β-**3**), 3.50 (*dt*, *J* = 7.3, 3.4 Hz, 2 H, H_2C(5'), α/β-**3**), 3.38 (*s*, 3 H, H_3CO, α-**3**), 3.28 (*s*, 3 H, H_3CO, β-**3**), 2.23-1.97 (*m*, 2 H, H_2C(2'), α/β-**3**).

^{13}C-NMR (75 MHz, $CDCl_3$): δ 138.2, 138.2, 138.1, 138.0 (arom. C_{ipso}, α/β-**3**), 128.4, 128.4, 127.9, 127.7, 127.7, 127.6 (arom. C, α/β-**3**), 105.5 (C(1), β-**3**), , 105.2 (C(1), α-**3**), 82.9 (C(4), β-**3**), 82.2 (C(4), α-**3**), 80.0 (C(3), β-**3**), 78.6 (C(3), α-**3**), 73.3 (C(5), β-**3**), 72.0 (C(5), α-**3**), 71.6 (benz. CH_2, β-**3**), 70.2 (benz. CH_2, α-**3**), 55.2 (CH_3O, β-**3**), 55.0 (CH_3O, α-**3**), 39.4 (C(2), β-**3**), 38.9 (C(2), α-**3**).

ESI-MS: 351.1 (100, $[M+Na]^+$).

5.2.3 1-(3', 5'-di-*O*-benzyl-2'-deoxy-α/β-D-ribofurano-1-yl)-propan-2-one (5)



3 (5.42 g, 16.50 mmol) in 200 ml dry MeCN were cooled to 0 °C under Ar and silylenolether **4** (3.88 g, 29.78 mmol) was added. Then, $SnCl_4$ (2.7 ml, 5.99 g, 22.99 mmol) was added dropwise. It was stirred 2 h at 0 °C. The solvents were evaporated and the residue was dissolved in CH_2Cl_2, washed twice with water and once with sat. $NaHCO_3$ solution. It was dried over Na_2SO_4 and purified by Column Chromatography (SiO_2, Hex/Et_2O 1:1). **5** (3.45 g, 59 %) was obtained as a yellow oil in a α/β 3:1 ratio. Treatment of the product with $Zn(OAc)_2$ (16.64 g, 90.73 mmol) and NaOMe (4.90 g, 90.73 mmol) in 150 ml MeOH for 3 d at rt shifted the anomeric ratio to α/β 1:2.

R_f (SiO_2, Hex/Et_2O 1:1): 0.15

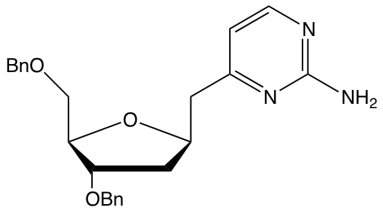
IR (film): 3062*w*, 3030*m*, 2891*w*, 2863*w*, 1712*s*, 1604*w*, 1496*w*, 1454*m*, 1359*m*, 1273*w*, 1205*w*, 1163*w*, 1093*s*, 1028*m*, 912*w*, 737*s*, 698*s*, 603*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.31-7.21 (*m*, 10 H, arom. H), 4.49-4.44 (*m*, 5 H, benz. H, HC(1'), α/β-**5**), 4.15 (*q*, *J* = 4.4 Hz, 1 H, HC(4'), α-**5**), 4.07 (*dt*, J = 4.9, 2.6 Hz, 1 H, HC(4'), β-**5**), 4.03 (*dd*, *J* = 3.4, 0.6 Hz, 1 H, HC(3'), α-**5**), 3.98 (*dt*, J = 6.2, 1.8 Hz, HC(3'), β-**5**), 3.49-3.36 (*m*, 2 H, H_2C(5'), α/β-**5**), 2.90 (*dd*, *J* = 16.5, 6.8 Hz, 1 H, HC(1), α-**5**), 2.74 (*dd*, *J* = 16.0, 6.9 Hz, 1 H, HC(1), β-**5**), 2.65 (*dd*, J = 16.4, 6.5 Hz, 1 H, HC(1), α-**5**), 2.54 *(dd*, *J* = 16.0, 5.8 Hz, 1 H, HC(1), β-**5**), 2.33 (*dt*, J = 13.1, 6.7, 1 H, HC(2'), α-**5**), 2.18 (*dd*, *J* = 5.2, 1.5 Hz, 1 H, HC(2'), β-**5**), 2.12 (*s*, H_3CO, β-**5**), 2.11 (*s*, H_3CO, α-**5**), 1.71 (*ddd*, J = 13.1, 5.6, 4.2, 1 H, HC(2'), α-**5**), 1.57 (*ddd*, *J* = 13.2, 10.2, 6.3 Hz, 1 H, HC(2'), β-**5**).

^{13}C NMR (75 MHz, $CDCl_3$): δ 208.8 (CO, α-**5**), 208.3 (CO, β-**5**), 139.5, 139.4 (arom. C_{ipso}, α/β-**5**), 129.7, 129.0, 128.9 (arom C, α/β-**5**), 84.8 (C(4'), β-**5**), 83.8 (C(4'), α-**5**) 82.3 (C(3'), β-**5**), 82.2 (C(3'), α-**5**), 76.2 (C(1'), α-**5**), 76.0 (C(1'), β-**5**), 74.7 (C(5'), α/β-**5**), 72.8, 72.4, 72.2, 72.1 (benz. CH_2, α/β-**5**), 51.2 (C(1), α-**5**), 50.6 (C(1), β-**5**), 39.5 (C(2'), β-**5**), 39.1 (C(2'), α-**5**), 31.9 (C(3), α-**5**), 32.0 (C(3), β-**5**).

ESI-MS: 377.26 (100, $[M+Na]^+$).

5.2.4 2-Amino-4-((3',5'-di-*O*-benzyl-2'-deoxy-β-D-ribofurano-1-yl)methyl)pyrimidine (7 β)



To **5** (0.20 g, 0.57 mmol) in 2 ml dry toluene, $BuOCH(NMe_2)_2$[135] (0.15 g, 0.86 mmol) was added under Ar. It was stirred for 20 h at 70 °C. After evaporation of the solvent, the residue was solved in 4 ml abs. EtOH under Ar. NaOEt (0.07 g, 1.08 mmol) was added and it was stirred for 4 h at 60 °C. After cooling down to 40 °C, guanidinium sulfate (0.47 g, 2.16 mmol) was added and it was stirred for 1.5 h at 40 °C. Then, it was stirred for 18 h at 70 °C. The solvents were evaporated. Purification by CC (SiO_2, Hex/EtOAc 1:8) gave a mixture of **7** and **8** (0.07 g, 22 %) as an anomeric mixture.

HPLC separation (Hex/MeOH/EtOH 97:2:1) provided **7**β as a yellow oil.

R_f (SiO_2, EtOAc/Hex 8:1): 0.25

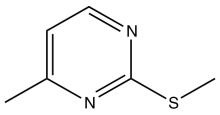
IR (film): 3333*s*, 3182*m*, 3089*m*, 3071*m*, 3033*m*, 2937*w*, 2867*s*, 1959*w*, 1892*w*, 1817*w*, 1636*s*, 1585*s*, 1496*m*, 1464*s*, 1368*m*, 1349*m*, 1276*m*, 12010*m*, 1099*s*, 1062*s*, 1034*w*, 916*w*, 813*w*, 748*s*, 712*s*, 618*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.16 (*d*, J = 5.0 Hz, HC(6)), 7.38–7.20 *(m*, arom. H), 6.57 (*d*, J = 5.1 Hz, HC(5)), 5.10 (*s br*, H_2N), 4.56 (*s*, $H_2COC(5')$), 4.57–4.45 (*m*, HC(1')), 4.48 (*s*, H2COC(3')), 4.17 (*dt*, J = 5.2, 2.6 Hz, HC(4')), 4.05 (dt, J = 6.6, 2.0, HC(3')); 3.54 (*dd*, J = 10.0, 3 J = 4.7 Hz, HC(5')), 3.44 (*dd*, J = 10.1, 3 J = 5.2 Hz, HC(5')), 2.90 (*dd*, J = 13.9, 6.9 Hz, HCC(4)), 2.77 (*dd*, J = 13.9, 5.7 Hz, HCC(4)), 2.12 (*ddd*, J = 13.2, 5.2, 1.5 Hz, $H_\alpha C(2')$), 1.72 (*ddd*, J = 13.2, 10.2, 6.4 Hz, $H_\beta C(2')$).

^{13}C NMR (75 MHz, $CDCl_3$): δ 168.4 (C(4)), 162.7 (C(2)), 157.9 (C(6)), 138.0, 137.9 (arom. C_{ipso}), 128.4, 128.43, 128.1, 127.7, 127.6, 127.5 (arom. C), 111.4 (C(5)), 83.5 (C(4')), 81.0 (C(3')), 77.2 (C(1')), 73.2 (COC(5')), 70.8 (COC(3')), 70.6 (C(5')), 43.2 (CC(4)), 38.0 (C(2')).

ESI-MS: 854.2 (100, $[M+Na]^+$).

5.2.5 4-methyl-2-(methylthio)-pyrimidine (15)



2-Mercapto-4-methylpyrimidine hydrochloride **14** (4.98 g, 30.63 mmol) was suspended in 60 ml EtOH and 60 ml 1 M aq. NaOH was added, followed by the addition of MeI (1.9 ml, 4.31 g, 30.39 mmol). It was stirred for 4 h at rt, then partially evaporated and extracted with EtOAc twice. The org. phase was washed with H_2O and sat. NaCl solution, dried over Na_2SO_4, evaporated under reduced pressure and dried at the HV overnight. **15** (3.50 g, 82 %) was obtained as a yellow liquid.

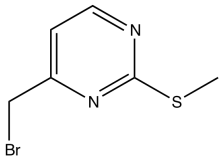
IR (film): 2926*w*, 1567*s*, 1541*s*, 1414*m*, 1372*w*, 1342*m*, 1325*s*, 1266*w*, 1217*s*, 1205*s*, 1182*m*, 1035*w*, 967*w*, 882*m*, 820*w*, 769*m*, 730*w*, 710*m*, 580*w*, 543*m*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.29 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 6.80 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 2.54 (*s*, 3 H, H_3CS), 2.44 (*s*, 3 H, H_3C)).

^{13}C NMR (75 MHz, $CDCl_3$): δ 172.2 (C(2)), 167.3 (C(4)), 156.6 (C(6)), 116.0 (C(5)), 24.1 (CH_3), 14.0 (SCH_3).

GC-MS: t_R = 5.20 min; 140 (100, M^+), 94 (76, M - SCH_3), 67 (74).

5.2.6 4-bromomehtyl-2-methylthiopyrimidine (16)



To **15** (0.5 ml, 0.50 g, 3.60 mmol) in 4.8 ml acetic acid, Br_2 (0.19 ml, 0.61 g, 3.81 mmol) was added by syringe and it was stirred for 2 h at 80 °C under N_2. After letting cool down to rt, Et_2O was added. The formed precipitation was filtered, treated with aq. $NaHCO_3$ solution and extracted with Et_2O. The org. phase was washed with sat. NaCl solution, dried over Na_2SO_4, and evaporated under reduced pressure. CC (SiO_2, CH_2Cl_2/hexane 1:1 to 100 % CH_2Cl_2) afforded **16** (0.43 g, 54 %) as a brown oil.

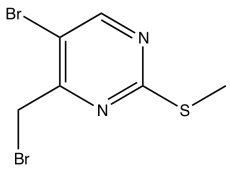
R_f (SiO_2, CH_2Cl_2/hexane 1:1): 0.39

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.52 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 7.11 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 4.36 (*s*, 2 H, H_2C), 2.57 (*s*, 3 H, H_3CS).

^{13}C-NMR (75 MHz, $CDCl_3$): δ 173.1 (C(2)), 165.3 (C(4)), 158.2 (C(6)), 115.4 (C(5)), 31.8 (CH_2Br), 14.4 (CH_3S).

GC-MS: t_R = 7.90 min; 220 (11, M^+), 218(10, M^+), 139 (100, M-Br), 93 (26), 66 (44).

5.2.7 5-bromo-4-bromomehtyl-2-methylthiopyrimidine (17)



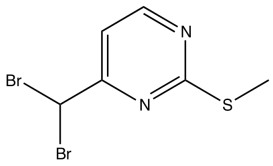
17 was obtained as a side product of the synthesis of **16** and isolated by CC purification (0.05 g, 5 %).

R_f (SiO_2, CH_2Cl_2/hexane 1:1): 0.81

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.56 (*s*, 1 H, HC(6)), 4.48 (*s*, 2 H, H_2C), 2.56 (*s*, 3H, H_3CS).

GC-MS: t_R = 9.62 min; 300 (7, M^+), 298 (13, M^+), 296 (7, M^+) 219 (87, M – Br), 217 (100, M – Br), 96 (45), 64 (74).

5.2.8 4-dibromomehtyl-2-methylthiopyrimidine (18)



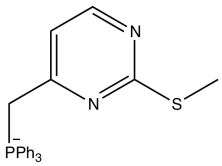
18 was obtained as a side product of the synthesis of **16** and isolated by CC purification (0.03 g, 3 %).

R_f (SiO_2, CH_2Cl_2/hexane 1:1): 0.51

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.62 (*d*, *J* = 5.2 Hz, 1 H, HC(6)), 7.39 (*d*, *J* = 5.2 Hz, 1 H, HC(5)), 6.40 (*s*, 1 H, $HCBr_2$), 2.58 (*s*, 3 H, H_3CS).

GC-MS: t_R = 9.36 min; 300 (8, M^+), 298 (14, M^+), 296 (8, M^+) 219 (89, M – Br), 217 (95, M – Br), 137 (89), 92 (63), 64 (100).

5.2.9 Triphenylphosphanylidene thiomethyl-pyrimidine (19)



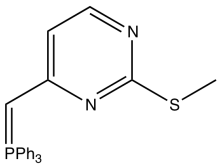
Triphenylphosphine (0.62 g, 2.37 mmol) in 2 ml dry benzene was added to **16** (0.37 g, 1.69 mmol) in 5 ml dry benzene and it was stirred for 4 h at rt under Ar. It was filtered, washed with benzene and dried at the HV overnight to give **19** (0.41 g, 60 %) as a bright pink solid.

M.p.: > 220 °C (decomposition).

IR (film): 3445*w*, 3048*w*, 2834*w*, 2339*w*, 1733*w*, 1556*s*, 1437*s*, 1350*m*, 1200*m*, 1109*s*, 996*w*, 913*w*, 868*w*, 748*m*, 720*s*, 688*s*, 511*s*, 444*w*.

^{1}H-NMR (300 MHz, DMSO-d_6): δ 8.43 (*d*, *J* = 5.0 Hz, 1 H, HC(6)), 7.77-7.63 (*m*, 15 H, Ph_3P), 7.05 (*d*, *J* = 5.0, 1 H, HC(5)), 5.40 (*d*, *J* = 15.6 Hz, 2 H, H_2C), 1.95 (*s*, 3 H, H_3CS).

5.2.10 Triphenylphosphine thiomethyl-pyrimidine (20)

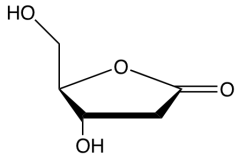


19 (3.00 g, 7.47 mmol) was solved in a biphasial mixture of 30 ml H_2O, 60 ml ice and 117 ml CH_2Cl_2. 13.4 ml 1 M aq. NaOH was added in one portion and it was stirred for 1 h at rt. The phases were separated and the water phase was extracted twice with CH_2Cl_2. The org. phase was dried over Na_2SO_4, evaporated under reduced pressure and dried at the HV overnight to give **20** (2.44 g, 82 %) as a beige foam.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.37 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 7.71-7.43 (*m*, 15 H, Ph_3P), 6.82 (*d*, *J* = 5.1, 1 H, HC(5)), 6.19 (*d*, *J* = 5.9 Hz, 1 H, HC=PPh_3), 2.56 (*s*, 3 H, H_3CS).

ESI-MS: 401 (100, $[M+H]^+$).

5.2.11 2-deoxy-D-ribono-1,4-lactone (21)



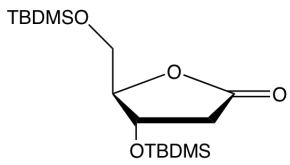
1 (7.12 g, 53.06 mmol) was dissolved in 230 ml H_2O and cooled to 0 °C. Br_2 was added dropwise and it was stirred for 16 h at 0 °C. It was evaporated under reduced pressure, the residue was dissolved in acetone, filtered through celite and washed with acetone. It was evaporated under reduced pressure and dried at the HV to give **21** (7.00 g, 99 %) as a dark viscous oil.

IR (film): 3415*w*, 1771*m*, 1190*w*, 1056*w*, 505*w*, 417*m*.

^{1}H-NMR (300 MHz, aceton-d_6): δ 4.38 (*dt*, *J* = 6.4, 2.2 Hz, 1 H, HC(4)), 4.22 (*m*, 1 H, HC(3)), 3.61 (*dd*, *J* = 4.1, 3.7 Hz, 2 H, $H_2C(5)$), 2.73 (*dd*, *J* = 17.7, 6.5 Hz, 1 H, HC(2)), 2.18 (*dd*, *J* = 17.7, 1.9 Hz, 1 H, HC(2)).

^{13}C NMR (75 MHz, aceton-d_6): δ 175.4 (C=O), 88.1 (C(4)), 68.5 (C(3)), 61.5 (C(5)), 37.9 (C(2)).

5.2.12 3,5-bis-*O*-(tert-butyldimethylsilyl)-2-deoxy-D-ribono-1,4-lactone (22)



To lactone **21** (7.01 g, 53.06 mmol) and imidazole (15.08 g, 221.46 mmol) in 60 ml dry DMF at 0 °C under Ar, TBDMSCl (17.50 g, 116.10 mmol) in 40 ml dry DMF was added dropwise. It was stirred for 21 h at rt under Ar. The reaction mixture was poured into 300 ml of EtOAc, filtered over celite and washed with EtOAc. The filtrate was partially evaporated, then

washed with 0.5 M aq. HCl, H_2O and sat. NaCl solution, dried over Na_2SO_4 and evaporated. Filtration through a 5 cm layer of SiO_2 (CH_2Cl_2-hexane 2:1) afforded **22** (12.14 g, 63 %) as a beige solid.

R_f (SiO_2, CH_2Cl_2/hexane 2:1): 0.29

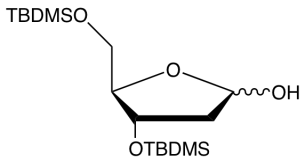
M.p.: 76.0 - 78.2 °C.

IR (KBr): 2957*s*, 2857*s*, 1769*s*, 1473*m*, 1390*m*, 1361*m*, 1260*s*, 1172*s*, 1127*m*, 1094*m*, 1008*m*, 969*m*, 944*m*, 916*w*, 839*s*, 781*s*, 665*m*, 601*w*, 540*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 4.50 (*dt*, *J* = 6.6, 2.4 Hz, 1 H HC(4)), 4.32 (*q*, *J* = 2.6 Hz, 1 H, HC(3)), 3.78 (*dd*, *J* = 3.2, 2.3 Hz, 2 H, $H_2C(5)$), 2.81 (*dd*, *J* = 17.7, 6.7 Hz, 1 H, HC(2)), 2.38 (*dd*, *J* = 17.6, 2.6 Hz, 1 H, HC(2)), 0.88 (*t*, *J* = 2.1 Hz, 18 H, 2x $(H_3C)_3CSi$), 0.08-0.06 (*m*, 12 H, 2x $(H_3C)_2Si$)).

^{13}C-NMR (75 MHz, $CDCl_3$): δ 175.7 (C=O), 88.6 (C(4)), 70.9 (C(3)), 63.5 (C(5)), 39.4 (C(2)), 26.27, 26.16 ($(H_3C)_3CSi$), 18.9, 18.6 ($(H_3C)_3CSi$), -4.6, -4.6, -5.3, -5.4 ($(CH_3)_2Si$).

5.2.13 3,5-bis-*O*-(t-butyldimethylsilyl)-2-deoxy-D-ribofuranose (23)



DIBAlH (1.5 M in toluene, 33.5 ml, 50.25 mmol) was added dropwise to **22** (12.14 g, 33.65 mmol) in 150 ml dry Et_2O at -78 °C under Ar and it was stirred for 5 h. 35 ml MeOH were added dropwise and it was let warm to 0 °C. Na/K-tartrate (0.5 M, 100 ml) was added and it was stirred overnight at rt. Et_2O was added, separated and the org. phase was washed with 0.5 M Na/K-tartrat solution and H_2O. It was dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, 100 % CH_2Cl_2 to CH_2Cl_2/ Et_2O 8:1) afforded **23** (11.61 g, 95 %) as a light yellow oil.

R_f (SiO_2, CH_2Cl_2/hexane 2:1): 0.15

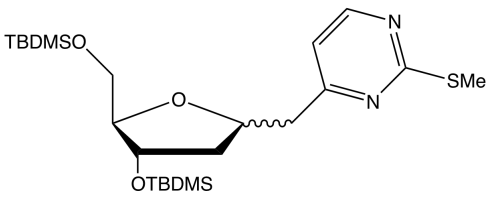
IR (film): 3428*w*, 2954*m*, 2929*m*, 2886*w*, 2858*m*, 1472*w*, 1362*w*, 1254*m*, 1118*m*, 1083*m*, 1006*w*, 836*s*, 776*s*, 675*w*.

^{1}H-NMR (300 MHz, DMSO-d_6): δ 6.17 (*d*, *J* = 5.3 Hz, 1 H, α-OH), 6.02 (*d*, *J* = 5.3 Hz, 1 H, β-OH), 5.33 (*q*, *J* = 4.7 Hz, 1 H, HC(1), α-**23**), 5.22 (*td*, *J* = 5.3, 3.7 Hz, 1 H, HC(1), β-**23**), 4.24 (*ddd*, *J* = 5.7, 4.7, 3.2 Hz, 1 H, HC(3)), 4.05 (*dt*, *J* = 7.9, 6.1 Hz, 1 H, HC(4)), 3.63-3.48 (*m*, 2 H, H_2C(5)), 2.28 (*ddd*, *J* = 13.2, 7.8, 5.5 Hz, 1 H, HC(2)), 1.53 (*ddd*, *J* = 13.1, 6.1, 3.6 Hz, 1 H, HC(2)), 0.83, 0.82, 0.81, 0.80 (4 *s*, 18 H, 2x $(H_3C)_3CSi$), 0.00, 0.00, -0.01, -0.02 (4 *s*, 12H, 2x $(H_3C)_2Si$).

^{13}C-NMR (75 MHz, DMSO-d_6): δ 97.7, 96.7 (C(4)), 85.8, 83.2 (C(3)), 72.8, 71.4 (C(1)), 64.3, 62.4 (C(5)), 42.5, 42.3 (C(2)), 25.7, 25.6 ($(H_3C)_3CSi$), 18.0, 17.6 ($(H_3C)_3CSi$), -4.8, -4.9, -5.0, -5.1 ($(CH_3)_2Si$).

ESI-MS: 385.3 (100, $[M+Na]^+$).

5.2.14 2-Thiomethyl-4-((3',5'-di-*O*-*t*-butyldimethylsilyl-2'-deoxy-α/β-D-ribofurano-1-yl)pyrimidine (25)



23 (1.86 g, 5.12 mmol) and **20** (2.44, 6.09 mmol) were refluxed in dry toluene for 36 h under N_2. It was evaporated and the residue was resolved in 30 ml MeOH. NaOMe (0.06 g, 1.17 mmol) was added and it was stirred for 2 h. 1 M NH_4Cl and EtOAc was added, separated and the org. phase was washed with sat. NaCl. solution, dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, Hexane/EtOAc 20:1 to 10:1) afforded **25** (1.79 g, 72 %) as a pale yellow oil with a α/β ratio of 1:2.

α- and β-anomers were separated by normal phase HPLC on a nitrile column (isocratic 0.1% iPrOH in hexane, 30 ml/min).

25-β:

R_f (SiO_2, hexane/EtOAc 5:1): 0.32

IR (film): 2953*m*, 2927*m*, 2898*w*, 2856*m*, 1568*m*, 1543*m*, 1463*w*, 1335*w*, 1254*m*, 1201*w*, 1201*w*, 1092*m*, 1006*w*, 938*w*, 835*s*, 777*s*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.34 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 6.88 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 4.48 (*td*, *J* = 10.6, 4.6 Hz, 1 H, HC(1')), 4.24-4.22 (*m*, 1 H, HC(3')), 3.74 (*ddd*, *J* = 5.7, 3.8, 2.1 Hz, 1 H, HC(4')), 3.46 (*ddd*, *J* = 53.3, 10.7, 4.9 Hz, 2 H, $H_2C(5')$), 2.89-2.84 (*m*, 2 H, $H_2CC(4)$), 2.50 (*s*, 3 H, H_3CS), 1.87-1.64 (*m*, 2 H, $H_2C(2')$), 0.83 (*d*, *J* = 7.0 Hz, 18 H, 2x $(H_3C)_3CSi$), 0.02, 0.00, -0.02, -0.03 (4 *s*, 12 H, 2x $(H_3C)_2Si$).

^{13}C-NMR (100 MHz, $CDCl_3$): δ 172.6 (C(4)), 168.7 (C(2)), 157.0 (C(6)), 116.9 (C(5)), 88.4 (C(4')), 77.4 (C(3')), 74.4 (C(1')), 64.3 (C(5')), 44.2 (CC(4)), 41.5 (C(2')), 26.48, 26.32 ((CH$_3$)$_3$CSi), 18.9, 18.5 ((CH$_3$)$_3$$C$Si), 14.6 ($CH_3$S), -4.14, -4.19, -4.77, -4.87 (CH$_3$)$_2$Si).

ESI-MS: 485.45 (100, $[M+H]^+$), 507.51 (5, $[M+Na]^+$).

HPLC (hexane/iPrOH 99:1): t_R = 9.4 min.

25-α:

R_f (SiO_2, hexane/EtOAc 5:1): 0.32

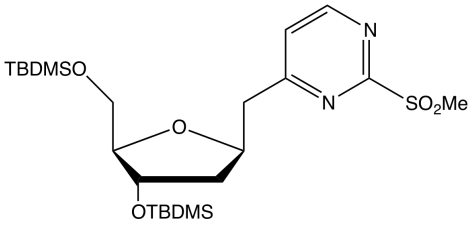
^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.32 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 6.85 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 4.44 (*ddd*, *J* = 7.4, 5.5, 2.0 Hz, 1 H, HC(1')), 4.29 (*dt*, *J* = 6.2, 3.3 Hz, 3 H, HC(3')), 3.83 (*dt*, *J* = 4.8, 3.7 Hz, 1 H, HC(4')), 3.55-3.40 (*m*, 2 H, $H_2C(5')$), 3.10-2.83 (*m*, 2 H, $H_2CC(4)$), 2.49 (*s*, 3 H, H_3CS)), 2.17 (*dt*, *J* = 12.9, 6.6 Hz, 1 H, $H_2C(2')$), 1.66 (*ddd*, *J* = 12.7, 4.9, 4.3 Hz, 1 H, $H_2C(2')$), 0.83 (*d*, *J* = 7.0 Hz, 18 H, 2x $(H_3C)_3CSi$), 0.01- -0.06 (*m*, 12 H, 2x $(H_3C)_2Si$).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.2 (C(4)), 168.7 (C(2)), 156.9 (C(6)), 116.8 (C(5)), 87.04 (C(4')), 78.0 (C(3')), 73.7 (C(1')), 63.7 (C(5')), 44.7 (CC(4)), 40.4 (C(2')), 26.10, 25.99 ((CH$_3$)$_3$CSi), 18.5, 18.1 ((CH$_3$)$_3$$C$Si), 14.3 ($CH_3$S), -4.53, -4.59, -5.18, -5.24 ((CH$_3$)$_2$Si).

ESI-MS: 485.45 (100, $[M+H]^+$), 507.50 (7, $[M+Na]^+$).

HPLC (hexane/iPrOH 99:1): t_R = 10.3 min.

5.2.15 3',5'-TBDMS-protected sulfonopyrimidine nucleoside (26)



To **25** (0.11 g, 0.23 mmol) in 1.5 ml dry CH_2Cl_2 under N_2, mCPBA (0.14 g, 0.82 mmol) in 2 ml dry CH_2Cl_2 was added and stirred for 16 h at rt. The white precipitation that formed overnight was filtered off and the filtrate was washed with 10 % $Na_2S_2O_3$ solution and H_2O, dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, Hexane/EtOAc 3:1) afforded **26** (0.09 g, 78%) as a colourless oil.

R_f (SiO_2, hexane/EtOAc 3:1): 0.14

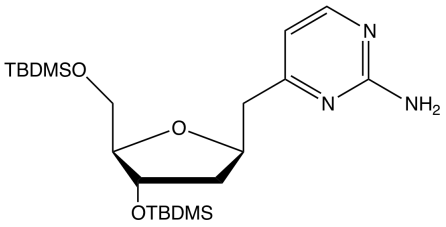
IR (film): 2953*m*, 2929*m*, 2894*m*, 2857*m*, 1579*m*, 1471*w*, 1360*w*, 1322*s*, 1254*s*, 1137*s*, 1091*s*, 1034*w*, 1007*w*, 961*w*, 837*s*, 778*s*, 667*w*, 540*m*, 419*m*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.77 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 7.57 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 4.56-4.50 (*m*, 1 H, HC(1')), 4.31 (*dd*, *J* = 3.7, 1.8 Hz, 1 H, HC(3')), 3.79 (*ddd*, *J* = 5.4, 3.7, 1.9 Hz, 1 H, HC(4')), 3.60 (*dd*, *J* = 10.8, 3.8 Hz, 2 H, H_2C(5')), 3.35 (*s*, 3 H, H_3CS) 3.19-2.98 (*m*, 2 H, HCC(4)), 1.98-1.71 (*m*, 2 H, H_2C(2')), 0.88 (*d*, *J* = 6.8 Hz, 18 H, $(H_3C)_3CSi$), 0.06 (*s*, 12 H, $(H_3C)_2Si$).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 171.0 (C(4)), 165.7 (C(2)), 157.8 (C(6)), 124.2 (C(5)), 88.1 (C(4')), 76.6 (C(3')), 73.7 (C(1')), 63.7 (C(5')), 43.7 (*C*C(4)), 41.2 (CH_3SO_2), 39.2 (C(2')), 26.0, 25.8 ((CH_3)$_3$CSi), 18.4, 18.0 ((CH_3)$_3$*C*Si), -4.6, -4.7, -5.3, -5.4 ((CH_3)$_2$Si).

ESI-MS: 517.48 (70, $[M+H]^+$), 539.51 (100, $[M+Na]^+$), 555.51 (75, $[M+K]^+$).

5.2.16 3',5'-TBDMS-protected 2-aminopyrimidine nucleoside (27)



26 (0.09 g, 0.18 mmol) was solved in 2 ml dioxane in a pressure tube. 25 % aq. NH_3 was added and it was stirred for 4 h at 80 °C, then evaporated under reduced pressure. CC (SiO_2, EtOAc) afforded **27α/β** (0.06 g, 77 %, α/β = 1:5) as a colourless oil.

27β:

R_f (SiO_2, EtOAc): 0.69

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.18 (*d*, *J* = 5.2 Hz, 1 H, HC(6)), 6.66 (*d*, *J* = 5.2 Hz, 1 H, HC(5)), 5.34 (*s br*, 2 H, H_2N), 4.54-4.47 (*m*, 1 H, HC(1')), 4.30 (*dt*, *J* = 5.5, 2.0 Hz, 1 H, HC(3')), 3.80 (*ddd*, *J* = 6.0, 3.9, 2.1 Hz, 1 H, HC(4')), 3.53 (*ddd*, *J* = 57.5, 10.7, 5.0 Hz, 2 H, H_2C(5')), 2.83-2.80 (*m*, 2 H, HCC(4)), 1.93-1.69 (*m*, 2 H, H_2C(2')), 0.89 (*d*, *J* = 6.4 Hz, 18 H, $(H_3C)_3CSi$)), 0.08-0.03 (*m*, 12 H, $(H_3C)_2Si$).

^{13}C-NMR (100 MHz, $CDCl_3$): δ 169.4 (C(6)), 161.8 (C(4)), 156.3 (C(2)), 111.4 (C(5)), 88.2 (C(4')), 77.5 (C(3')), 74.0 (C(1')), 64.0 (C(5')), 43.5 (*C*C(4)), 41.3 (C(2')), 26.1, 26.0 ((CH_3)$_3$C(CH_3)$_2$Si), 18.6, 18.2 ((CH_3)$_3$*C*(CH_3)$_2$Si), -4.5, -4.4 -5.2, -5.1 ((CH_3)$_3$C(*C*H_3)$_2$Si).

ESI-MS: 454.38 (100, $[M+H]^+$).

27α:

R_f (SiO_2, EtOAc): 0.69

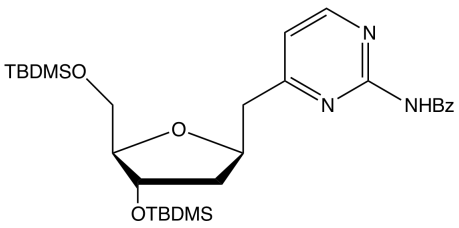
^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.18 (*d*, *J* = 5.2 Hz, 1 H, HC(6)), 6.64 (*d*, *J* = 5.2 Hz, 1 H, HC(5)), 5.43 (*s br*, 2 H, H_2N), 4.39-4.37 (*m*, 1 H, HC(1')), 4.30 (*dt*, *J* = 5.5, 2.0 Hz, 1 H, HC(3')), 3.91 (*ddd*, *J* = 3.3, 3.0, 2.2 Hz, 1 H, HC(4')), 3.56 (*dq*, *J* = 27.3, 4.6 Hz, 2 H,

$H_2C(5')$), 3.11-2.78 (*m*, 2 H, HCC(4)), 2.29-1.68 (*m*, 2 H, $H_2C(2')$), 0.89 (*d*, *J* = 6.4 Hz, 18 H, $(H_3C)_3CSi$)), 0.08-0.03 (*m*, 12 H, $(H_3C)_2Si$).

^{13}C-NMR (100 MHz, $CDCl_3$): δ 170.4 (C(6)), 161.8 (C(4)), 156.3 (C(2)), 111.6 (C(5)), 87.2 (C(4')), 77.8 (C(3')), 73.7 (C(1')), 63.8, (C(5')), 44.3 (*C*C(4)), 40.6 (C(2')), 26.1, 26.0 ((CH_3)$_3$C(CH_3)$_2$Si), 18.6, 18.2 ((CH_3)$_3$*C*(CH_3)$_2$Si), -4.5, -4.6, -5.1, -5.2 ((CH_3)$_3$C(*C*H_3)$_2$Si).

ESI-MS: 454.38 (100, $[M+H]^+$).

5.2.17 3',5'-TBDMS-protected benzamine-pyrimidine nucleoside (28)



BzCl (0.03 g, 0.22 mmol) was added dropwise to **27** (0.02, 0.05 mmol) in 1 ml dry pyridine at 0 °C under Ar, then it was stirred for 2 h at rt. It was cooled to 0 °C, then 0.08 ml H_2O were added and it was stirred for 15 min at 0 °C. 0.27 ml 25 % aq. NH_3 solution were added and it was stirred for another 30 min at 0 °C. It was evaporated and the residue was dissolved in Et_2O. It was washed twice with water and the water phases were extracted once with Et_2O. The org. phase was dried over Na_2SO_4 and evaporated under reduced pressure. Purification by CC (SiO_2, hexane/EtOAc 1:1) gave **28** (0.17 g, 61 %) as a colourless oil.

R_f (SiO_2, hexane/EtOAc): 0.48

IR (film): 3246*w*, 3060*w*, 2927*m*, 2856*m*, 1753*m*, 1716*m*, 1593*m*, 1563*m*, 1532*m*, 1450*m*, 1389*w*, 1253*m*, 1184*m*, 1087*m*, 835*s*, 776*s*, 737*s*, 670*w*.

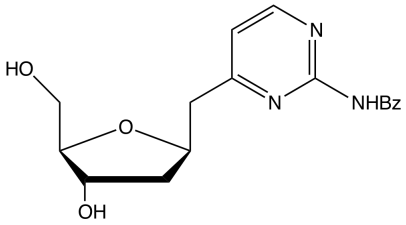
^{1}H-NMR (300 MHz, $CDCl_3$): δ 9.59 (*s br*, 1 H, HN), 8.62 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 8.14-8.02 (*m*, 2 H, arom. H_{meta}), 7.64-7.45 (*m*, 3 H, arom. $H_{ortho,\ para}$), 7.13 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 4.56 (*ddt*, *J* = 9.9, 7.5, 5.0 Hz, 1 H, α/β-HC(1'), 4.41-4.37 (*m*, 4.41-4.37, 1 H, α-HC(3')), 4.38-4.30 (*m*, 1 H, β-HC(3')), 3.98-3.94 (*m*, 1 H, α-HC(4')) 3.82 (*ddd*, *J* = 5.9, 3.9,

2.0 Hz, 1 H, β-HC(4')), 3.66-3.42 (*m*, 2 H, α/β-H_2C(5')), 3.06-2.93 (*m*, 2 H, α/β-HCC(4)), 1.98-1.73 (*m*, 2 H, α/β-H_2C(2')), 0.92-0.87 (*m*, 18 H, $(H_3C)_3CSi$)), 0.09-0.03 (*m*, 12 H, $(H_3C)_2Si$).

^{13}C-NMR (100 MHz, $CDCl_3$): *δ* 169.7 (C(6), α/β-**28**), 165.0 (C=O, α/β-**28**), 158.3 (C(4), α/β-**28**), 157.7 (C(2), α/β-**28**), 134.7, 132.5, 129.0, 127.7 (arom Bz-CH, α/β-**28**), 117.2 (C(5), α/β-**61**), 88.2 (C(4'), β-**28**), 87.1 (C(4'), α-**28**), 77.9 (C(3'), β-**28**), 77.4 (C(3'), α-**28**), 74.0 (C(1'), β-**28**), 73.7 (C(1'), α-**28**), 64.0 (C(5'), β-**28**), 63.8 (C(5'), α-**28**), 44.7 (*C*C(4), α-**28**), 43.9 (*C*C(4), β-**28**), 41.3 (C(2'), β-**28**), 40.5 (C(2'), α-**28**), 26.2, 26.1, 26.0 ($(CH_3)_3C(CH_3)_2Si$, α/β-**28**), 18.6, 18.5, 18.2, 18.2 ($(CH_3)_3C(CH_3)_2Si$, α/β-**28**), -4.4, -4.5, -4.6, -5.1, -5.2, -5.2 ($(CH_3)_3C(CH_3)_2Si$, α/β-**28**).

ESI-MS: 558.63 (100, $[M+H]^+$).

5.2.18 *N*-(4-((2'-deoxy-β-D-ribofurano-1-yl)-methyl)pyrimidin-2-yl)benzamide (29)



28 (0.02 g, 0.03 mmol) and NH_4F in 1 ml MeOH were refluxed for 15 h, then the solvent was evaporated. EtOAc and H_2O were added, separated and the water phase was extracted with EtOAc. The water phase was evaporated under reduced pressure and the residue purified by CC (SiO_2, EtOAc/MeOH 10:1) to give **29** (0.01 g, 70 %).

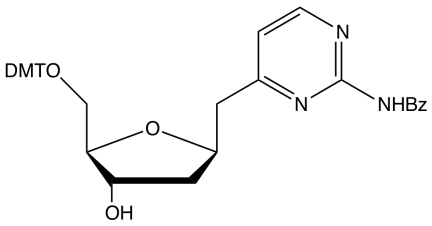
R_f (SiO_2, EtOAc/MeOH 10:1): 0.11

^{1}H-NMR (500 MHz, MeOD): *δ* 8.56 (*d*, *J* = 5.1 Hz, 1 H, HC(6)), 8.01-7.98 (*m*, 2 H, arom. H_{meta}), 7.62-7.42 (*m*, 3 H, arom. $H_{ortho,\ para}$), 7.21 (*d*, *J* = 5.2 Hz, 1 H, HC(5)), 4.62-4.56 (*m*, 1 H, HC(1')), 4.23 (*dt*, *J* = 5.9, 2.4 Hz, 1 H, HC(3')), 3.79 (*td*, *J* = 4.8, 2.7 Hz, 1 H, HC(4')), 3.53 (*dd*, *J* = 4.8, 1.6 Hz, 2 H, H_2C(5')), 3.03-3.02 (*m*, 2 H, HCC(4)), 2.02-1.85 (*m*, 2 H, H_2C(2')).

^{13}C-NMR (125 MHz, MeOD): δ 170.4 (C(6)), 167.6 (C=O), 158.3 (C(4)), 158.1 (C(2)), 132.7, 129.8, 128.9, 128.28 (arom. C (Bz)), 117.7 (C(5)), 88.2 C(4')), 77.6 (C(3')), 73.1 (C(1')), 63.1 (C(5')), 43.3 (*C*C(4)), 40.8 (C(2')).

ESI-MS: 330.16 (100, $[M+H]^+$), 352.16 (70, $[M+Na]^+$).

5.2.19 *N*-(4-((5'-*O*-(4'',4'''-dimethoxytrytil)-2'-deoxy-β-D-ribofurano-1-yl)-methyl)pyrimidin-2-yl)benzamide (30)



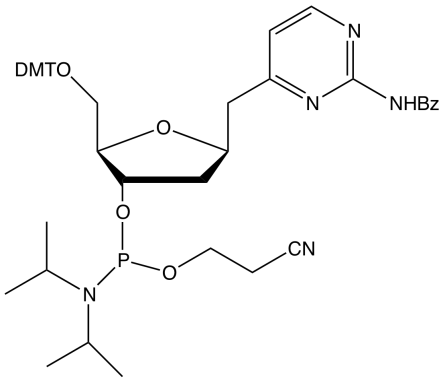
29 (0.01 g, 0.02 mmol), DMTCl (0.01 g, 0.04 mmol), DMAP (0.001 g, 0.008 mmol) and $NEt(iPr)_2$ (0.01 ml, 0.008 g, 0.06 mmol) in 0.5 ml dry pyridine were stirred for 24 h at rt under Ar. It was evaporated and the residue was dissolved in CH_2Cl_2 containing 1 % TEA, washed twice with H_2O, dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, EtOAc/TEA 200:1) afforded **30** (0.005 g, 36 %) as a yellow oil.

R_f (SiO_2, EtOAc/TEA 200:1): 0.47

^{1}H-NMR (500 MHz, $CDCl_3$): δ 8.56 (*d*, *J* = 5.1 Hz, 1 H. HC(6)), 7.90-7.88 (*m*, 2 H, benz. H), 7.58-7.17 (*m*, 12 H, 3 arom. H (Bz) and 9 arom. H (Tr)), 7.02 (*d*, *J* = 5.1 Hz, 1 H, HC(5)), 6.82-6.80 (*m*, 4 H, arom. H (Tr)), 4.62 (*dd*, *J* = 9.4, 6.3 Hz, 1 H, HC(1')), 4.34-4.33 (*m*, 1 H, HC(3')), 3.93 (*ddd*, *J* = 5.4, 4.4, 3.0 Hz, 1 H, HC(4')), 3.78 (*s*, 3 H, H_3CO), 3.15 (*ddd*, *J* = 59.9, 9.7, 5.1 Hz, 2 H, $H_2C(5')$), 3.03-2.98 (*m*, 2 H, HCC(4)), 2.07-1.92 (*m*, 2 H, $H_2C(2')$).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 169.2 (C(4)), 164.8(C=O), 158.5 (C_{para} (PhOMe)), 157.5 (C(6)), 157.3 (C(2)), 144.9 (C_{ipso} (Ph)), 136.1 (C_{ipso} (PhOMe)), 134.4 (C_{ipso} (Bz)), 132.2 (C_{para} (Bz)) 130.1 (C_{ortho} (PhOMe)), 128.8, 128.2, 127.9, 127.5 (C_{ortho} (Ph), C_{meta} (Ph), C_{ortho} (Bz), C_{meta} (Bz)), 126.8 (C_{para} (Ph)), 117.0 (C(5)), 113.1 (C_{meta} (PhOMe)), 86.0 (C(4'), 84.6 (C_{tert} (Tr)), 77.3 (C(3')), 74.2 (C(1')), 64.4 (C(5')), 55.3 (CH_3O), 43.7 (*C*C(4)), 40.6 (C(2')).

5.2.20 (1-((N$^{2'}$-Benzoyl(2-aminopyrimidin-4-yl))methyl)-5-*O*-(4,4-dimethoxytrityl)-2-deoxy-β-D-ribofurano-1-yl)-β-cyanoethyl-*N*,*N*-diisopropyl phosphoramidite (31)



To **30** (0.005 g, 0.007 mmol) in 0.05 ml dry CH_2Cl_2 under Ar, $NEt(^iPr)_2$ (0.005 ml, 0.004 g, 0.03 mmol) was added by syringe, followed by $PClN(^iPr)_2OEtCN$ (0.01 g, 0.04 mmol) in 0.05 ml dry CH_2Cl_2. It was stirred for 3 h at rt, then 0.2 ml MeOH was added, followed by EtOAc containing 0.5 % of TEA. The org. phase was washed with H_2O, sat. $NaHCO_3$ solution and sat. NaCl solution, dried over Na_2SO_4 and evaporated under reduced pressure. Purification by CC (SiO_2, CH_2Cl_2/EtOAc/TEA 100:100:1) gave **31** (0.006 g, 91 %) as a pale yellow foam.

R_f (SiO_2, CH_2Cl_2/EtOAc/TEA 100:100:1): 0.82

^{1}H-NMR (500 MHz, $CDCl_3$): δ 8.58 (*dd*, *J* = 5.1, 1.6 Hz, 1 H, HC(6)), 7.90 (*dd*, *J* = 6.7, 1.3 Hz, 2 H, HC_{ortho} (Bz)), 7.56-7.52 (*m*, 1 H, HC_{para} (Bz)), 7.49-7.43 (*m*, 2 H, H_{meta} (Bz)), 7.35-7.16 (*m*, 9 H, arom. H (Tr)), 7.06 (*d*, *J* = 4.4 Hz, 1 H, HC(5)), 6.83-6.78 (*m*, 2 H, HC_{meta} (PhOMe)), 4.61-4.56 (*m*, 1 H, HC(1')), 4.46-4.42 (*m*, 1 H, HC(3')), 4.21 (*td*, *J* = 9.8, 6.0 Hz, 1 H, HC(4')), 3.77, 3.75 (2 *s*, 6 H, H_3CO), 3.66-3.64 (*m*, 2 H, $H_2C(5')$), 3.57 (*q*, *J* = 7.3 Hz, 2 H, OCH_2CH_2CN), 3.31-3.03 (*m*, 2 H, $NCHMe_2$), 2.93 (*q*, *J* = 7.3 Hz, 2 H, $H_2CC(4)$), 2.61-2.55 (*m*, 2 H, OCH_2CH_2CN), 2.10-1.85 (*m*, 2 H, $H_2C(2')$), 1.27-1.21 (*m*, 12 H, $((CH_3)_2CH)_2N$).

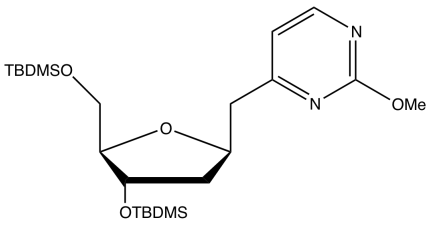
^{13}C-NMR (125 MHz, $CDCl_3$): δ 169.1 (C(4)), 164.6 (C=O), 158.4 (C_{para} (PhOMe)), 157.5 (C(6)), 157.2 (C(2)), 144.9 (C_{ipso} (Ph)), 136.0 (C_{ipso} (PhOMe)), 134.6 (C_{ipso} (Bz)), 132.0 (C_{para} (Bz)), 130.1 (C_{ortho} (PhOMe)), 128.7, 128.2, 127.8, 127.3 (C_{ortho} (Ph), C_{meta} (Ph), C_{ortho} (Bz), C_{meta} (Bz)),126.7 (C_{para} (Ph)), 117.6 (CN), 116.8 (C(5)), 113.1 (C_{meta}, (PhOMe), 86.0 (C_{tert} (Tr)), 84.9 (C(4')), 76.9 (C(1')), 75.0 (C(3')), 64.1 (C(5')), 58.1 (OCH_2CH_2CN), 55.2

(CH_3O), 43.8 (*C*C(4)), 43.1 (N*C*HMe$_2$), 38.9 (C(2')), 24.7 ((CH_3)$_2$CH)$_2$N), 20.0 (OCH_2CH_2CN).

^{31}P-NMR (203 MHz, $CDCl_3$): δ 147.9, 147.8.

ESI-MS: 832.61 (18, $[M+H]^+$), 303.19 (100, DMT).

5.2.21 3',5'-TBDMS-protected methoxypyrimidine nucleoside (35)



Sulfone **26** (0.07 g, 0.14 mmol) and NaOMe (0.05 g, 0.89 mmol) in 5 ml MeOH were refluxed for 3 h under N_2. After cooling down to rt, CH_2Cl_2 was added and the org. phase was washed twice with H_2O, dried over Na_2SO_4 and evaporated under reduced pressure. It was purified by CC (SiO_2, hexane/EtOAc 2:1) which gave **35** (0.05 g, 78 %) as a colourless oil.

R_f (SiO_2, hexane/EtOAc 2:1): 0.80

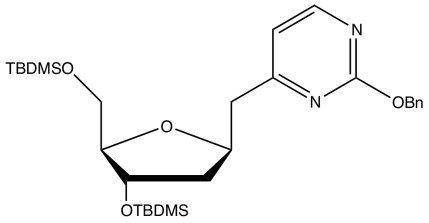
^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.39 (*d*, *J* = 4.9 Hz, 1 H, HC(6), **35**-α/β), 6.94-6.89 (*m*, 1 H, HC(5), **35**-α/β), 4.56-4.53 (*m*, 1 H, HC(1'), **35**-α/β), 4.45-4.35 (*m*, 1 H, HC(3'), **35**-α), 4.30-4.28 (*m*, 1 H, HC(3'), **35**-β), 4.01 (*s*, 3 H, H_3CO, **35**-α/β), 3.91-3.90 (*m*, 1 H, HC(4'), **35**-α), 3.80-3.78 (*m*, 1 H, HC(4'), **35**-β), 3.63-3.40 (*m*, 2 H, $H_2C(5')$, **35**-α/β), 3.20-2.95 (*m*, 2 H, $H_2CC(4)$, **35**-α), 2.92 (*d*, *J* = 6.3 Hz, 2 H, $H_2CC(4)$, **35**-β), 2.30-1.71 (*m*, 2 H, $H_2C(2')$, **35**-α/β), 0.88 (*d*, *J* = 7.7 Hz, 18 H, $(H_3C)_3CSi$, **35**-α/β), 0.07-0.02 (*m*, 12 H, $(H_3C)_2Si$, **35**-α/β).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.6 (C(4), **35**-β), 172.3 (C(4), **35**-α), 168.7 (C(2), **35**-β), 168.6 (C(2), **35**-α), 157.1 (C(6), **35**-β), 156.9 (C(6), **35**-α), 116.9 (C(5), **35**-β), 116.7 (C(5), **35**-α), 88.6 (C(4'), **35**-β), 88.1 (C(4'), **35**-α), 77.6 (C(3'), **35**-α), 77.3 (C(3'), **35**-β), 73.9 (C(1'), **35**-α), 73.7 (C(1'), **35**-β), 64.2 (C(5'), **35**-α), 64.0 (C(5'), **35**-β), 54.9 (OCH_3, **35**-α), 54.4 (OCH_3, **35**-β), 44.7 (*C*C(4), **35**-α), 44.1 (*C*C(4), **35**-β), 41.4 (C(2'), **35**-β), 40.7 (C(2'), **35**-α), 26.5, 26.2 (((*C*H$_3$)$_3$CSi), **35**-β), 26.0, 25.9 ((*C*H$_3$)$_3$CSi), **35**-α), 19.1, 18.6 ((CH$_3$)$_3$*C*Si),

35-β), 18.6, 18.0 ((CH$_3$)$_3$*C*Si), -4.1, -4.1, -4.7, -4.8 (*C*H$_3$)$_2$Si, **35**-β), -4.5, -4.6, -5.1, -5.2 ((*C*H$_3$)$_2$Si, **35**-β).

ESI-MS: 469.45 (100, [M+H]$^+$).

5.2.22 3',5'-TBDMS-protected benzoxypyrimidine nucleoside (37)

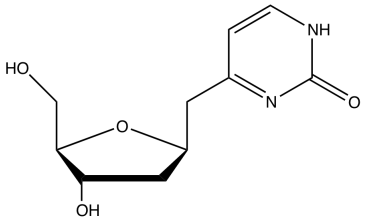


To **26** (0.07 g, 0.13 mmol) in 5 ml dry benzyl alcohol under Ar, NaOBn (1 M in benzyl alcohol, 0.75 ml, 0.75 mmol) was added dropwise by syringe. It was stirred for 2.5 h at 70 °C, then CH_2Cl_2 was added. The org. phase was washed with H_2O, dried over Na_2SO_4 and evaporated under reduced pressure. It was dried at the HV at 60 °C overnight. CC (SiO_2, hexane/EtOAc 3:1) gave **37** (0.04 g, 58 %) as a colourless oil.

R_f (SiO_2, hexane/EtOAc 3:1): 0.33

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.40 (*d*, *J* = 5.0 Hz, 1 H, HC(6), **37**-α/β), 7.51-7.30 (*m*, 5 H, arom. H (Bn) **37**-α/β), 6.93 (*d*, *J* = 5.0 Hz, 1 H, HC(5), **37**-β), 6.90 (*d*, *J* = 5.0 Hz, 1 H, **37**-α), 5.44 (*s*, 2 H, H_2COC(2), **37**-α/β), 4.61-4.52 (*m*, 1 H, HC(1'), **37**-α/β), 4.36 (*dt*, *J* = 6.1, 3.6 Hz, HC(3'), **37**-α), 4.29 (*dt*, *J* = 5.4, 1.9 Hz, 1 H, **37**-β), 3.92 (*dt*, *J* = 5.1, 3.6 Hz, 1 H, HC(4'), **37**-α), 3.64 (*ddd*, *J* = 5.9, 3.9, 2.0 Hz, 1 H, HC(4'), **37**-β), 3.64-3.39 (*m*, 2 H, H_2C(5'), **37**-α/β), 3.16 (*dd*, *J* = 13.7, 7.9 Hz, 2 H, HCC(4), **37**-α), 2.92 (*m* , 2 H, HCC(4), **37**-β), 2.27-1.68 (*m*, 2 H, H_2C(2'), **37**-α/β), 0.91, 0.88 (2 *d*, *J* = 1.7 Hz, 18 H, (H_3C)$_3$CSi, **37**-α/β), 0.08-0.03 (*m*, 12 H, (H_3C)$_2$Si, **37**-α/β).

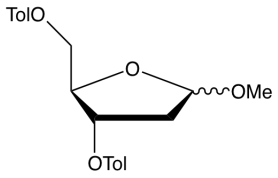
5.2.23 4-((2'-deoxy-α/β-D-ribofurano-1-yl)methyl)pyrimidin-2-one (36)



37 (0.04 g, 0.08 mmol) was solved in 4.8 ml dry CH_2Cl_2 under Ar and it was cooled to 0 °C. BBr_3 (1 M in CH_2Cl_2, 0.45 ml, 0.45 mmol) was added dropwise and it was stirred for 3 min at 0 °C. 3 ml H_2O were added, followed by 1 ml sat. $NaHCO_3$ solution. It was separated and the water phase was evaporated under reduced pressure. Purification by CC (SiO_2, EtOAc/MeOH 1:1) afforded **36** (0.004 g, 24 %) as a light yellow solid.

ESI-MS: 249.22 (75, $[M+Na]^+$).

5.2.24 2-deoxy-1-*O*-methyl-3,5-di-*O*-*p*-toluoyl-D-ribofuranose (38)



2 (2.65 g, 17.87 mmol) was solved in 20 ml dry pyridine under Ar and cooled to 0 °C. *p*-Toluoyl chloride (5.6 ml, 6.11 g, 39.49 mmol) was added dropwise by syringe. It was stirred overnight at RT. The reaction mixture was diluted with cold water and extracted 3 times with Et_2O. The combined org. layers were washed with sat. $NaHCO_3$ solution, H_2O and sat. NaCl solution. It was dried over Na_2SO_4 and evaporated. CC (SiO_2, Hex/EtOAc 2:1) afforded **38** (5.72 g, 83 %) as a white solid in an anomeric mixture.

R_f (SiO_2, Hex/EtOAc 3:1): 0.49

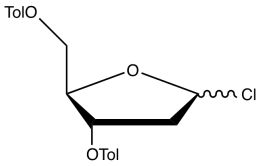
IR (film): 2952*w*, 2925*w*, 1718*m*, 1611*w*, 1445*w*, 1376*w*, 1270*m*, 1209*w*, 1177*w*, 1104*w*, 1067*w*, 1020*w*, 753*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.92-7.83 (*m*, 4 H, Ar-H), 7.19-7.13 (*m*, 4 H, Ar-H), 5.55-5.32 (*m*, 1 H, HC(3)), 5.15 (*td*, *J* = 7.5, 2.8 Hz, 1 H, HC(1)), 4.58-4.38 (*m*, 3 H, HC(5) and HC(4)), 3.42, 3.36 (2 *s*, 2 H, α/β-OCH_3), 2.49 (*ddd*, *J* = 14.1, 7.3, 2.2 Hz, 1 H, HC(2)), 2.41, 2.39 (2 *s*, 2 H, α/β-CH_3), 2.12 (*ddd*, *J* = 14.6, 2.2, 0.9 Hz, 1 H, HC(2)).

^{13}C NMR (75 MHz, $CDCl_3$): δ 166.3 (CO, α-**38**), 166.1 (CO, β-**38**), 144.0, 143.7 (C_{ipso} (Tol)), 129.8, 129.7, 129.1, 129.1 (C_{ortho}, C_{meta} (Tol)), 127.2, 126.9 (C_{para} (Tol)), 105.6 (C(1), β-**38**), 105.1 (C(1), α-**38**), 81.9 (C(4), β-**38**), 81.0 (C(4), α-**38**), 75.4 (C(3), β-**38**), 74.6 (C(3), α-**38**), 65.2 (C(5), β-**38**), 64.3 (C(5), α-**38**), 55.2 (OCH_3, β-**38**), 55.1 (OCH_3, α-**38**), 39.3 (C(2), α/β-**38**), 21.7 (CH_3 (Tol), α/β-**38**).

ESI-MS: 407.2 (100, $[M+Na]^+$).

5.2.25 Toluoylprotected chlorosugar (39)



A HCl solution was prepared by dropwise addition of acetyl chloride (0.88 ml, 0.96 g, 12.27 mmol) to 4.05 ml acetic acid and 0.2 ml H_2O at 0 °C. 4.5 ml of this HCl solution was added dropwise to **38** (1.71 g, 4.43 mmol) in 3.5 ml acetic acid at rt. Then, it was cooled to 0 °C and an additional amount of acetyl chloride (0.2 ml, 0.22 g, 2.80 mmol) was added. The formed precipitate was filtered and washed with cold Et_2O. Drying at the HV overnight gave **39** (1.30 g, 75 %) as a white solid.

M.p.: 108.2 - 108.8 °C.

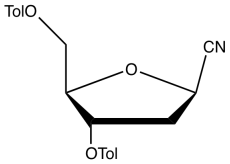
IR (film): 1718*s*, 1682*s*, 1612*m*, 1577*w*, 1420*m*, 1377*w*, 1273*s*, 1178*m*, 1107*m*, 1020*w*, 959*w*, 839*w*, 753*s*, 690*w*, 465*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.99 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.90 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.28-7.23 (*m*, 4 H, arom. H), 6.48 (*d*, *J* = 4.9 Hz, 1 H, HC(1)), 5.52 (*ddd*, *J* = 7.3,

2.9, 1.2 Hz, 1 H, HC(3)), 4.86 (*q*, *J* = 3.6 Hz, 1 H, HC(4)), 4.74-4.61 (*m*, 2 H, $H_2C(5)$), 2.94-2.76 (*m*, 2 H, $H_2C(2)$), 2.43, 2.42 (2 *s*, 6 H, H_3CPh).

^{13}C-NMR (75 MHz, $CDCl_3$): δ 166.4, 166.4 (C=O), 144.3, 144.1 (C_{ipso} (Tol)), 129.9, 129.7, 129.3, 129.2 (C_{ortho}, C_{meta} (Tol)), 127.5, 127.3 (C_{para} (Tol)), 95.4 (C(4)), 84.8 (C(3)), 72.9 (C(1)), 63.7 (C(5)), 44.2 (C(2)), 21.7, 21.6 (CH_3 (Tol)).

5.2.26 Toluoylprotected cyanoribose (40)



To **39** (3.60 g, 9.25 mmol) in 50 ml dry CH_2Cl_2 under Ar, TMS-CN (1.26 g, 12.68 mmol) and BF_3-Et_2O (2.38 g, 16.78 mmol) were added and it was stirred overnight at rt. 50 ml sat. $NaHCO_3$ solution was added, it was separated and the water phase was extracted twice with EtOAc. The org. phases were dried over Na_2SO_4, evaporated under reduced pressure and purified by CC (SiO_2, hexane/EtOAc 6:1 to 3:1) to give pure β-anomer **40** (1.87 g, 53 %) as a white solid and **40α** (0.39 g, 11 %) as a side product.

40β

R_f (SiO_2, hexane/EtOAc/ 6:1): 0.49

M.p.: 140.4 - 143.0 °C.

IR: 1715*s*, 1611*m*, 1450*w*, 1409*w*, 1379*w*, 1266*s*, 1177*m*, 1102*s*, 1019*m*, 979*w*, 841*w*, 751*s*, 690*w*, 573*w*.

^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.97 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.89 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.26-7.24 (*m*, 4 H, arom. H), 5.60 (*d*, *J* = 5.8 Hz, 1 H, HC(3)), 4.91 (*dd*, *J* = 9.1, 6.7 Hz, 1 H, HC(1)), 4.62-4.52 (*m*, 3 H, HC(4), $H_2C(5)$), 2.77-2.61 (*m*, 2 H, $H_2C(2)$), 2.42, 2.41 (2 *s*, 6 H, H_3C (Tol)).

^{13}C-NMR (100 MHz, $CDCl_3$): δ 166.4, 165.9 (C=O), 144.8, 144.4 (C_{ipso} (Tol)), 130.0, 129.9, 129.5, 129.5 (C_{ortho}, C_{meta} (Tol)), 126.9, 126.5 (C_{para} (Tol)), 118.1 (CN), 84.1 (C(4)), 75.6 (C(3)), 66.1 (C(1)), 64.0 (C(5)), 38.1 (C(2)), 21.9, 21.9 (CH_3 (Tol)).

ESI-MS: 380.26 (100, $[M+H]^+$).

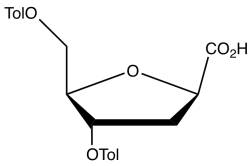
40α:

R_f (SiO_2, hexane/EtOAc/ 6:1): 0.46

^{1}H-NMR (400 MHz, $CDCl_3$): δ 8.04 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.94 (*d*, *J* = 8.2 Hz, 2 H, arom. H), 7.31-7.27 (*m*, 4 H, arom. H), 5.66-5.64 (*m*, 1 H, HC(3)), 5.08 (*dd*, *J* = 8.2, 1.6 Hz, 1 H, HC(1)), 4.71 (*td*, *J* = 4.1, 2.0 Hz, 1 H, HC(4)), 4.57 (*qd*, *J* = 11.9, 4.1 Hz, 2 H, H_2C(5)), 2.73 (*ddd*, *J* = 14.3, 8.2, 6.0 Hz, 1 H, HC(2)), 2.63 (*dt*, *J* = 14.3, 1.4 Hz, 1 H, HC(2)), 2.45, 2.44 (2 *s*, 6 H, H_3C (Tol)).

^{13}C-NMR (100 MHz, $CDCl_3$): δ 166.25, 166.23 (C=O), 144.7, 144.4 (C_{ipso} (Tol)), 130.2, 129.9, 129.53, 129.47 (C_{ortho}, C_{meta} (Tol)), 126.9, 126.5 (C_{para} (Tol), 118.7 (CN), 84.5 (C(4)), 75.1 (C(3)), 66.7 (C(1)), 63.8 (C(5)), 38.1 (C(2)), 21.9, 21.9 (CH_3 (Tol)).

5.2.27 Toluoylprotected carboxylic acid ribose (41)



To **40β** (1.18 g, 3.11 mmol) in 25 ml 1,4-dioxane was added 2.7 ml of 32% HCl and it was refluxed for 6 h. H_2O was added and it was extracted twice with CH_2Cl_2 dried over $MgSO_4$ and evaporated. It was continued without further purification.

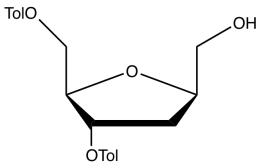
IR (film): 3192*w*, 2952*w*, 1716*s*, 1611*m*, 1445*w*, 1376*w*, 1269*s*, 1177*m*, 1101*s*, 1020*m*, 840*w*, 752*s*, 690*w*, 477*w*.

^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.95-7.91 (*m*, 4 H, arom. H), 7.28-7.23 (*m*, 4 H, arom. H), 5.52 (*dd*, *J* = 4.4, 1.5 Hz, 1 H, HC(3)), 4.82 (*dd*, *J* = 9.9, 6.6 Hz, 1 H, HC(1)), 4.65 (*dd*, *J* = 12.7, 6.0 Hz, 1 H, HC(4)), 4.60-4.55 (*m*, 2 H, $H_2C(5)$), 2.66 *(ddd*, *J* = 14.0, 6.7, 1.8 Hz, 1 H, $H_2C(2)$), 2.48-2.45 (*m*, 1 H, $H_2C(2)$), 2.43, 2.40 (2 *s*, H_3C (Tol)).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 173.0 (COOH), 167.2, 166.4 (C=O), 144.8, 144.6 (C_{ipso} (Tol)), 130.1, 130.0, 129.5, 129.5 (C_{ortho}, C_{meta} (Tol)), 126.8, 126.6 (C_{para} (Tol)), 84.9 (C(4)), 75.7 (C(3)), 67.3 (C(1)), 64.6 (C(5)), 36.6 (C(2)), 22.0, 21.9 (CH_3 (Tol)).

ESI-MS: 397.26 (100, $[M-H]^-$).

5.2.28 Toluoylprotected methylhydroxy ribose(42)



To crude **13** (1.48 g, 3.71 mmol) in 30 ml dry THF, $BH_3(Me)_2S$-complex (2 M, 3.0 ml, 5.79 mmol) was added at 0 °C. It was stirred for 16 h at rt under Ar. H_2O was added and it was extracted twice with CH_2Cl_2. The org. phase was washed with H_2O and sat. NaCl solution, dried over $MgSO_4$ and evaporated. CC (SiO_2, hexane/EtOAc 2:1) afforded **42** (0.69 g, 58 % over two steps) as a colourless oil.

R_f (SiO_2, hexane/EtOAc 2:1): 0.24

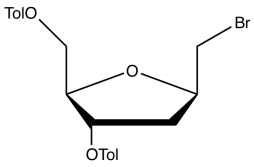
IR (film): 3484*w*, 2921*w*, 2877*w*, 1715*m*, 1611*w*, 1377*w*, 1269*s*, 1177*m*, 1105*s*, 1020*m*, 960*w*, 841*w*, 752*s*, 690*w*.

^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.94-7.89 (*m*, 4 H, arom. H), 7.26-7.22 (*m*, 4 H, arom. H), 5.50 (*d*, *J* = 6.3 Hz, 1 H, HC(3)), 4.63-4.59 (*m*, 1 H, HC(1)), 4.49-4.42 (*m*, 3 H, HC(4) and $H_2C(5)$), 3.91 (*dd*, *J* = 12.2 Hz, 2.8 Hz, 1 H, H_2COH), 3.56 (*dd*, *J* = 12.2, 3.2 Hz, 1 H, H_2COH), 2.42-2.37 (*m*, 1 H, $H_2C(2)$), 2.42, 2.41 (2 *s*, 6 H, H_3C (Tol)), 2.14 (*dd*, *J* = 5.6, 1.2 Hz, 1 H, $H_2C(2)$).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 167.0, 166.4 (C=O), 144.3, 144.2 (C_{ipso} (Tol)), 130.0, 129.9, 129.4, 129.3 (C_{ortho}, C_{meta} (Tol)), 127.2, 127.1 (C_{para} (Tol)), 83.3 (C(4)); 80.2 (C(3)); 77.3 (C(1)); 64.9 (C(5)); 63.2 (CH_2OH); 33.4 (C(2)); 21.9, 21.8 (CH_3 (Tol)).

ESI-MS: 385.16 (40, $[M+H]^+$), 407.13 (100, $[M+Na]^+$).

5.2.29 Toluoylprotected bromomethyl ribose (43)



CBr_4 (2.13 g, 6.43 mmol) and PPh_3 (1.96 g, 7.47 mmol) were added to **42** (0.69 g, 1.79 mmol) in 10 ml dry CH_2Cl_2 at 0 °C under Ar. It was stirred for 1 h at rt. Then, the solvents were evaporated under reduced pressure. It was purified by CC (SiO_2, hexane/EtOAc 1:1) to give **43** (0.74 g, 93 %) as a viscous oil.

R_f (SiO_2, hexane/EtOAc 1:1): 0.67

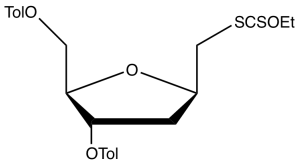
IR (film): 2952*w*, 2922*w*, 1714*s*, 1611*m*, 1442*w*, 1408*w*, 1375*w*, 1266*s*, 1176*m*, 1099*s*, 1020*m*, 923*w*, 840*w*, 751*s*, 690*w*, 476*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.94, 7.93 (2 *d*, *J* = 6.2 Hz, 4 H, arom. H), 7.26-7.23 (*m*, 4 H, arom. H), 5.55-5.53 (*m*, 1 H, HC(3)), 4.52-4.47 (*m*, 4 H, HC(1), HC(4), $H_2C(5)$), 3.53-3.49 (*m*, 2 H, CH_2Br), 2.42, 2.41 (2 *s*, 6 H, H_3C Tol)), 2.35 (*ddd*, *J* = 13.8, 5.4, 1.4 Hz, 1 H, $H_2C(2)$), 2.20 (*ddd*, *J* = 13.8, 9.8, 6.2 Hz, 1 H, $H_2C(2')$).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 166.5, 166.3 (C=O), 144.3, 144.1 (C_{ipso} (Tol)), 130.1, 129.5 (C_{ortho}, C_{meta} (Tol)), 127.3, 127.1 (C_{para} (Tol)), 83.6 (C(4)), 78.4 (C(3)), 76.7 (C(1)), 64.8 (C(5)), 37.7 (CH_2Br), 34.5 (C(2)), 21.9, 21.8 (CH_3 (Tol)).

ESI-MS: 447.19 (98, $[M+H]^+$), 449.19 (100, $[M+H]^+$).

5.2.30 Toluoylprotected methylxanthate ribose (44)



To **43** (0.06 g, 0.13 mmol) in 2 ml dry acetone under Ar, potassium ethyl xanthogenate (0.06 g, 0.39 mmol) was added and it was refluxed for 2 h. The solvent was eyaporated and it was dissolved in CH_2Cl_2. The organic layer was washed with water, sat. NaCl solution and dried over Na_2SO_4. CC (SiO_2, hexane to hexane/EtOAc 3:1) afforded **44** (0.04 g, 60 %) as a colourless oil.

R_f (SiO_2, hexane/EtOAc 1:1): 0.66

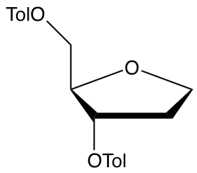
IR (film): 2954*w*, 2921*w*, 1715*s*, 1611*m*, 1508*w*, 1448*w*, 1408*w*, 1374*w*, 1266*s*, 1212*m*, 1176*m*, 1100*s*, 1045*s*, 1019*s*, 922*w*, 840*m*, 751*s*, 690*m*, 607*w*.

^{1}H-NMR (500 MHz, $CDCl_3$): δ 7.96-7.91 (*m*, 4 H, arom. H), 7.26-7.23 (*m*, 4 H, arom. H), 5.50 (*d*, *J* = 6.1 Hz, 1 H, HC(3)), 4.66-4.62 (*m*, HC(1)), 4.55-4.39 (*m*, 3 H, HC(4), $H_2C(5)$), 3.55-3.47 (*m*, H_2CS), 2.42, 2.41 (2 *s*, 6 H, H_3C (Tol)), 2.38-2.09 (*m*, 2 H, $H_2C(2)$), 1.41 (*t*, *J* = 7.1 Hz, 3 H, H_3CH_2CO), 1.28-1.24 (*m*, 2 H, H_3CH_2CO).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 214.7 (CSO), 166.5, 166.2, (C=O), 144.4, 144.1 (C_{ipso} (Tol)); 130.0, 129.4 (C_{ortho}, C_{meta} (Tol)), 127.3, 127.1 (C_{para} (Tol)), 83.3 (C(4)), 77.4 (C(3)), 70.6 (C(1)), 64.6 (C(5)), 40.1 (CH_3CH_2O), 37.7 (CH_2SCSO), 34.5 (C(2)), 21.9, 21.9 (CH_3 Tol)), 14.0 (CH_3CH_2O).

ESI-MS: 489.37 (100, $[M+H]^+$), 511.45 (7, $[M+Na]^+$).

5.2.31 1,2-dideoxy-3,5-di-O-(p-toluyl)-d-ribofuranose (47)



To a solution of **38** (3.15 g, 8.20 mmol) in 20 ml dry CH_2Cl_2 at 0 °C under Ar, triethylsilane (2.9 ml, 2.12 g, 18.21 mmol) and BF_3-etherate (2.3 ml, 2.60 g, 18.31 mmol) were added by syringe. It was stirred for 1 h at 0 °C, then it was let warm to rt and stirred for 30 min. Sat. $NaHCO_3$ solution was added and it was stirred vigorously for 10 min. H_2O and CH_2Cl_2 were added and it was separated. The water phase was washed twice with CH_2Cl_2. The combined organic layers were dried over Na_2SO_4, filtered and the solvents were evaporated under reduced pressure. CC (SiO_2, Hex/Et_2O 1:1) afforded **47** (2.50 g, 86 %) as a bright yellow oil.

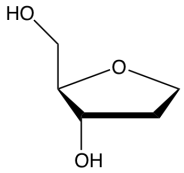
R_f (SiO_2, Hex/Et_2O 1:1): 0.47

IR (film): 2952*w*, 1715*s*, 1611*m*, 1508*w*, 1454*w*, 1408*w*, 1376*w*, 1268*s*, 1209*w*, 1177*m*, 1099*s*, 1020*w*, 840*w*, 752*s*, 690*w*.

^{1}H-NMR (300 MHz, CDCl3): δ 7.95-7.91 (*m*, 4 H, arom. H), 7.26-7.21 (*m*, 4 H, arom. H), 5.47 (*td*, *J* = 4.3, 2.1 Hz, 1 H, HC(3)), 4.49 (*td*, *J* = 4.3, 2.1 Hz, 2 H, HC(5)), 4.37 (*td*, *J* = 4.6, 2.4 Hz, 1 H, HC(4)), 4.20-4.00 (*m*, 2 H, HC(1)), 2.41 (2 *s*, 6 H, 2 H_3C), 2.37-2.15 (*m*, 2 H, HC(2)).

^{13}C NMR (75 MHz, $CDCl_3$): δ 166.35, 166.22 (CO), 144.0, 143.8 (C_{ipso} (Tol)), 129.8, 129.7, 129.1, 129.1 (C_{ortho}, C_{meta} (Tol)), 127.1, 127.0 (C_{para} (Tol)), 82.2 (C(4)), 76.5 (C(3)), 67.8 (C(1)), 64.6 (C(5)), 32.9 (C(2)), 21.7, 21.6 (CH_3 (Tol)).

5.2.32 1,2-dideoxy-D-ribofuranose (48)



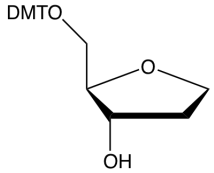
NaOMe (0.03 g, 0.58 mmol) in 10 ml MeOH was added to **47** (0.17 g, 0.48 mmol) under Ar and it was stirred for 7 h at RT. The solvent was evaporated and it was purified by CC (SiO_2, EtOAc/MeOH 20:1) to obtain **48** (0.07 g, 80 %).

R_f (SiO_2, CH_2Cl_2/MeOH 9:1): 0.33

^{1}H-NMR (300 MHz, DMSO-d_6): δ 4.86 (*d*, *J* = 4.1 Hz, 1 H, HOC(3)), 4.62 (*t*, *J* = 5.7 Hz, 1 H, HOC(5)), 4.06-4.02 (*m*, 1 H, HC(3)), 3.78-3.71 (*m*, 2 H, HC(1)), 3.56 (*td*, *J* = 5.3, 2.9 Hz, 1 H, HC(4)), 3.35-3.29 (*m*, 2 H, HC(5)), 1.94-1.85 (*m*, 1 H, HC(2)), 1.72-1.65 (*m*, 1 H, HC(2)).

^{13}C-NMR (75.5 MHz, MeOH-d_4): δ 88.4 (C(4)), 74.1 (C(3)), 68.6 (C(1)), 64.1 (C(5)), 36.0 (C(2)).

5.2.33 1,2-dideoxy-5-*O*-(4,4'-dimethoxytrityl)-D-ribofuranose (49)



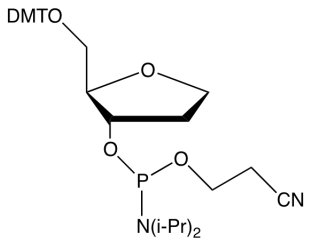
4-(dimethylamino)pyridine (0.01, 0.09 mmol) and DMTCl (0.31 g, 0.90 mmol) were both dried at the HV overnight, then solved in 5 ml dry pyridine under Ar. $NEt(iPr)_2$ (0.13 ml, 0.10 g, 0.75 mmol) and **48** (0.06 g, 0.50 mmol) in 1 ml dry pyridine were added. It was stirred for 20 h at RT under Ar. The solvent was evaporated and the residue was resolved in CH_2Cl_2 containing 1 % TEA. It was washed with sat. $NaHCO_3$ solution, H_2O and sat. NaCl solution, dried over Na_2SO_4 and evaporated. CC (SiO_2, CH_2Cl_2/EtOAc/TEA 75:14:1) afforded **49** (0.06 g, 44 %).

R_f (SiO_2, CH_2Cl_2/EtOAc/TEA 75:14:1): 0.37

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.44-6.81 (*m*, 13 H, arom. H), δ 4.31-4.27 (*m*, 1 H, HC(3)), 3.97 (*dd*, *J* = 8.3, 5.6 Hz, 2 H, $H_2C(1)$), 3.90-3.87 (*m*, 1 H, HC(4)), 3.80, 3.79 (2 *s*, 6 H, H_3CO), 3.25-3.10 (*m*, 2 H, $H_2C(5)$), 2.15-1.86 (*m*, 2 H, $H_2C(2)$).

^{13}C-NMR (75.5 MHz, $CDCl_3$): δ 158.5 (C_{para} (PhOMe)), 146.8 (C_{ipso} (Ph)), 136.0 (C_{ipso} (PhOMe)), 130.1 (C_{ortho} (PhOMe)), 128.1, 127.8 (C_{ortho} (Ph), C_{meta} (Ph)), 126.8 (C_{para} (Ph)), 113.1 (C_{meta}, (PhOMe), 86.1 (C_{tert} (Tr)), 85.0 (C(4)), 74.5 (C(3)), 67.1 (C(1)), 64.9 (C(5)), 55.2 (CH_3O), 34.9 (C(2)).

5.2.34 (1,2-dideoxy-5-*O*-(4,4'-dimethoxytrityl)-D-ribofuranoyl)-β-cyanoethyl-*N*,*N*-diisopropyl phosphoramidite (50)



To **49** (0.05 g, 0.12 mmol) in 0.3 ml dry CH_2Cl_2 under Ar, $PCl(N(iPr)_2OEtCN)$) (0.05 g, 0.20 mmol) in 1 ml dry pyridine and $NEt(iPr)_2$ (0.07 ml, 0.05 g, 0.41 mmol) were added. It was stirred for 2 h at RT, then MeOH was added. EtOAc containing 1 % TEA was added and it was washed with sat. $NaHCO_3$ solution, H_2O and sat. NaCl solution, dried over Na_2SO_4 and evaporated. CC (SiO_2, CH_2Cl_2/EtOAc/TEA 45:45:10) afforded **50** (0.06 g, 83 %) as a viscous oil.

R_f (SiO_2, CH_2Cl_2/EtOAc/TEA 45:45:10): 0.83

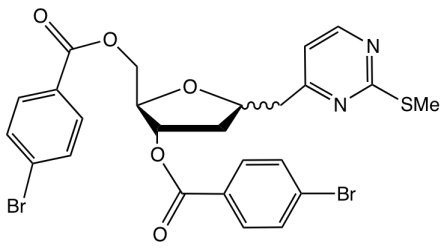
IR (film): 2965*w*, 1607*m*, 1508*s*, 1463*w*, 1395*w*, 1363*w*, 1300*w*, 1249*s*, 1176*m*, 1155*w*, 1061*m*, 1033*s*, 975*m*, 914*m*, 876*w*, 828*m*, 791*w*, 727*m*, 703*m*, 644*w*, 584*m*, 525*w*.

^{1}H-NMR (300 MHz, $CDCl_3$): δ 7.46-7.42 (*m*, 1 H, HC_{para} (Ph)), 7.35-7.19 (*m*, 8 H, HC_{ortho} (PhOMe), HC_{ortho} (Ph), HC_{meta} (Ph)), 6.84-6.80 (*m*, 4 H, HC_{meta} (PhOMe)), 4.45-4.36 (*m*, 1 H, HC(3)), 4.07-4.00 (*m*, 3 H $H_2C(1)$, HC(4)), 3.79, 3.78 (2 *s*, 6 H, H_3CO), 3.70-3.53 (*m*, 4 H, $H_2C(5)$, OCH_2CH_2CN), 3.16-3.09 (*m*, 2 H, $NCHMe_2$), 2.59, 2.44 (2 *t*, *J* = 6.5 Hz, 2 H, OCH_2CH_2CN), 2.17-1.93 (*m*, 2 H, $H_2C(2)$), 1.28-1.06 (*m*, 12 H, $((CH_3)_2CH)_2N$).

^{13}C-NMR (75.5 MHz, $CDCl_3$): δ 158.4 (C_{para} (PhOMe)), 144.9 (C_{ipso} (Ph)), 136.1 (C_{ipso} (PhOMe)), 130.1 (C_{ortho} (PhOMe)), 128.3, 127.8 (C_{ortho} (Ph), C_{meta} (Ph)), 126.7 (C_{para} (Ph)), 117.6 (CN), 113.1 (C_{meta}, (PhOMe), 86.0 (C_{tert} (Tr)), 84.9 (C(4)), 74.6 (C(3)), 67.5 (C(1)), 64.2 (C(5)), 58.3 (OCH_2CH_2CN), 55.2 (CH_3O), 43.2 ($NCHMe_2$), 34.4 (C(2)), 24.6, 24.5 ($((CH_3)_2CH)_2N$), 20.2 (OCH_2CH_2CN).

^{31}P-NMR (161 MHz, $CDCl_3$): δ 149.1, 148.8.

5.2.35 3',5'-*p*-Bromobenzoyl protected 2-thiomethylpyrimidine nucleoside (52)



To 4-bromobenzoyl chloride (0.48 g, 2.13 mmol) at 0 °C, **51** (0.60 g, 2.30 mmol) in 2.2 ml dry pyridine was added under Ar. It was stirred for 16 h at rt, then EtOAc was added and it was washed once with 5 % HCl solution in H_2O and twice with H_2O. It was dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, hexane/EtOAc 2:1) afforded **52** (0.28 g, 52 %) as a bright yellow waxlike solid.

R_f (SiO_2, Hexan/EtOAc 2:1): 0.52

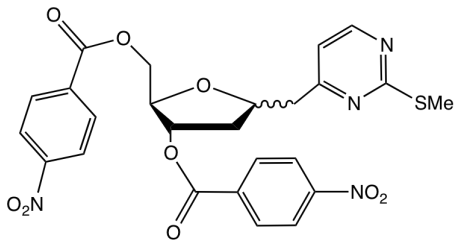
^{1}H-NMR (500 MHz, $CDCl_3$): δ 8.41 (*d*, *J* = 5.0 Hz, 1 H, HC(6), α-**52**), 8.38 (*d*, *J* = 5.0 Hz, 1 H, HC(6), β-**52**), 7.89 (*d*, 4 H, *J* = 8.5 Hz, arom. H_{ortho}, α/β-**52**), 7.60 (*dd*, *J* = 8.5, 2.5 Hz, 4 H, arom. H_{meta}, α/β-**52**), 6.91 (*d*, *J* = 5.0 Hz, 1 H, HC(5), α/β-**52**), 5.51-5.48 (*m*, 1 H, HC(1'), α/β-**52**), 4.74 (t, *J* = 6.5 Hz, 1 H, HC(3'), α-**52**), 4.64 (*dt*, *J* = 10.9, 5.4 Hz, 1 H, HC(3'), β-**52**), 4.56-4.46 (*m*, 2 H, $H_2C(5')$, α/β-**52**), 4.35-4.33 (*m*, 1 H, HC(4'), α/β-**52**), 3.15 (*d*, *J* = 7.2 Hz, 1 H, HCC(4), α-**52**), 3.00 (*dd*, *J* = 5.7, 2.2 Hz, 2 H, $H_2CC(4)$, β-**52**), 2.67 (*d*, *J* = 7.0 Hz, 1 H, HCC(4), α-**52**), 2.56 (*s*, 3 H, H_3CS, β-**52**), 2.54 (*s*, 3 H, H_3CS, α-**52**), 2.34 (*dd*, *J* = 13.9, 4.9 Hz, 1 H, HC(2'), α/β-**52**), 2.15 (*dt*, *J* = 9.3, 4.5 Hz, 1 H, HC(2'), α-**52**), 2.07 (*ddd*, *J* = 13.9, 10.6, 6.4 Hz, 1 H, HC(2'), β-**52**).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.5 (C(2), α-**52**), 172.4 (C(2), β-**52**), 167.0, 166.8 (CO, α/β-**52**), 165.6 (C(4), α-**52**), 165.5 (C(4), β-**52**), 165.4 (arom. C_{ipso}, α-**52**), 165.3 (arom. C_{ipso}, β-**52**), 157.0 (C(6), α-**52**), 156.9 (C(6), β-**52**), 131.9, 131.8, 131.2, 131.1 (C_{ortho}, C_{meta} (Tol), α/β-**52**), 128.7, 128.6 (C_{para} (Tol), α-**52**), 128.5, 128.4 (C_{para} (Tol), β-**52**), 116.6 (C(5), β-**52**), 116.3 (C(5), α-**52**), 82.5 (C(4'), β-**52**), 81.6 (C(4'), α-**52**), 77.8 (C(3'), α-**52**), 77.7 (C(3'), β-**52**), 77.2 (C(1'), α-**52**), 77.1 (C(1'), β-**52**), 64.8 (C(5'), β-**52**), 64.7 (C(5'), α-**52**), 43.7

(*C*C(4), α-**52**), 42.9 (*C*C(4), β-**52**), 38.3 (C(2'), β-**52**), 37.2 (C(2'), α-**52**), 14.1 (SCH_3, β-**52**), 14.1 (SCH_3, α-**52**).

ESI-MS: 645.26 (100, $[M+Na]^+$).

5.2.36 3',5'-*p*-Nitrobenzoyl protected 2-thiomethylpyrimidine nucleoside (53)



To 4-nitrobenzoyl chloride (0.07 g, 0.52 mmol) in 0.4 ml dry pyridine at 0 °C, **51** (0.06 g, 0.25 mmol) in 0.4 ml dry pyridine was added under Ar. It was stirred for 16 h at rt, then EtOAc was added and it was washed once with 5 % HCl solution in H_2O and twice with H_2O. It was dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, hexane/EtOAc 2:1) afforded **53** (0.07 g, 51 %) as a colourless viscous oil.

R_f (SiO_2, Hexan/EtOAc 2:1): 0.23

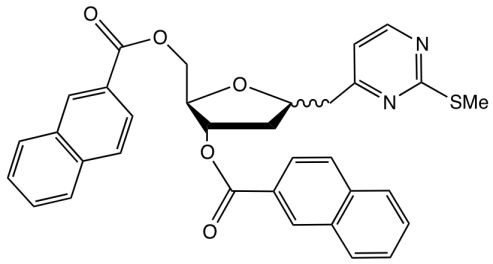
^{1}H-NMR (500 MHz, $CDCl_3$): δ 8.35 (*d*, *J* = 5.0 Hz, 1 H, HC(6), α-**53**), 8.31 (*d*, *J* = 5.0 Hz, 1 H, HC(6), β-**53**), 8.26-8.11 (*m*, 8 H, arom. H, α/β-**53**), 6.84 (*d*, *J* = 5.0 Hz, 1 H, HC(5), α/β-**53**), 5.45 (*d*, *J* = 6.4 Hz, 1 H, HC(1'), α/β-**53**), 4.71-4.64 (*m*, 1 H, HC(3), α-**53**), 4.60 (*dt*, *J* = 10.9, 5.5 Hz, 1 H, HC(3'), β-**53**), 4.52-4.43 (*m*, 2 H, H_2C(5'), α/β-**53**), 4.29 (*td*, *J* = 4.6, 2.1 Hz, 1 H, HC(4'), α/β-**53**), 3.08 (*dd*, *J* = 14.0, 7.1 Hz, 1 H, HCC(4), α-**53**), 2.97-2.94 (*m*, 2 H, H_2CC(4), β-**53**), 2.63 (*dd*, *J* = 14.1, 7.1 Hz, 1 H, HCC(4), α-**53**), 2.47 (*s*, 3 H, H_3C, β-**53**), 2.46 (*s*, 3 H, H_3C, α-**53**), 2.32 (*dd*, *J* = 14.2, 5.0 Hz, 1 H, HC(2'), α/β-**53**), 2.17-2.10 (m, 1 H, HC(2), α-**53**), 2.03 (*ddd*, *J* = 14.0, 10.7, 6.4 Hz, 1 H, HC(2), β-**53**).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.5 (C(2), α-**53**), 172.4 (C(2), β-**53**), 166.7, 166.6 (CO, α/β-**53**), 164. 3 (C(4), β-**53**), 164.2 (C(4), α-**53**), 157.0 (C(6), α-**53**), 156.9 (C(6), β-**53**), 150.9, 150.8 (arom. C_{para}, α/β-**53**), 135.1, 134.8 (arom. C_{ipso}, α/β-**53**), 130.9, 130.8 (arom. C_{ortho}, α/β-**53**), 123.6, 123.7 (arom. C_{meta}, α/β-**53**), 116.6 (C(5), β-**53**), 116.3 (C(5), α-**53**),

82.4 (C(4'), β-**53**), 81.5 (C(4'), α-**53**), 77.9 (C(3'), α-**53**), 77.8 (C(3'), β-**53**), 77.4 (C(1'), α-**53**), 77.2 (C(1'), β-**53**), 65.3 (C(5'), α/β-**53**), 43.5 (*C*C(4), α-**53**), 42.7 (*C*C(4), β-**53**), 38.1 (C(2'), β-**53**), 37.1 (C(2'), α-**53**), 14.1 (SCH_3, β-**53**), 14.1 (SCH_3, α-**53**).

ESI-MS: 555.54 (100, $[M+H]^+$).

5.2.37 3',5'-Naphtoyl protected 2-thiomethylpyrimidine nucleoside (54)



To 2-naphtoyl chloride (0.29 g, 1.52 mmol) in 1 ml dry pyridine at 0 °C, **51** (0.16 g, 0.61 mmol) in 1 ml dry pyridine was added under Ar. It was stirred for 16 h at rt, then EtOAc was added and it was washed once with 5 % HCl solution in H_2O and twice with H_2O. It was dried over Na_2SO_4 and evaporated under reduced pressure. CC (SiO_2, hexane/EtOAc 5:1) afforded **54** (0.05 g, 15 %) as a waxlike solid.

R_f (SiO_2, Hexan/EtOAc 2:1): 0.55

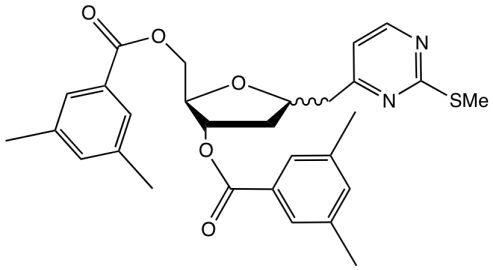
^{1}H-NMR (500 MHz, $CDCl_3$): δ 9.08 (*d*, *J* = 8.7 Hz, 1 H, HC(6), α/β-**54**), 8.96-7.48 (m, 14 H, arom. H, α/β-**54**), 6.91 (*d*, *J* = 5.0 Hz, 1 H, HC(5), α/β-**54**), 5.64 (*d*, *J* = 6.1 Hz, 1 H, HC(1'), β-**54**), 5.48-5.47 (m, 1 H, HC(1'), α-**54**), 4.82-4.80 (m, 1 H, HC(3'), α-**54**), 4.72-4.64 (m, 3 H, HC(3'), β-**54** and $H_2C(5')$, α/β-**54**), 4.51 (*dd*, *J* = 6.3, 4.2 Hz, 1 H, HC(4'), α/β-**54**), 3.18 (*dd*, *J* = 7.2, 6.8 Hz, HCC(4), α-**54**), 3.05-2.98 (m, 2 H, $H_2CC(4)$, β-**54**), 2.76 (*dd*, *J* = 7.2, 6.8 Hz, HCC(4), α-**54**), 2.53 (*s*, 3 H, H_3CS, β-**54**), 2.50 (*s*, 3 H, H_3CS, α-**54**), 2.40 (*dd*, *J* = 13.9, 4.9 Hz, 1 H, HC(2'), α/β-**54**), 2.18-2.12 (m, 1 H, HC(2'), α/β-**54**).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.4 (C(2), β-**54**), 171.8 (C(2), α-**54**), 167.2, 167.2 (CO, α/β-**54**), 167.1 (C(4), α-**54**), 167.0 (C(4), β-**54**), 157.0 (C(6), α-**54**), 156.9 (C(6), β-**54**), 134.5, 133.9, 133.7, 131.7, 130.5, 130.3, 128.7, 128.7, 128.6, 128.1, 128.0, 127.9, 126.4, 126.3, 126.0, 125.8, 125.6, 124.6, 124.5, 124.4 (arom. C (Naphtoyl), α/β-**54**), 116.6 (C(5), β-

54), 116.4 (C(5), α-**54**), 82.8 (C(4'), β-**54**), 81.9 (C(4'), α-**54**), 77.9 (C(3'), β-**54**), 77.7 (C(3'), α-**54**), 77.1 (C(1'), α-**54**), 77.0 (C(1'), β-**54**), 64.9 (C(5'), β-**54**), 64.8 (C(5'), α-**54**), 43.8 (*C*C(4), α-**54**), 43.1 (*C*C(4), β-**54**), 38.6 (C(2'), β-**54**), 37.5 (C(2'), α-**54**), 14.1 (SCH_3, β-**54**), 14.0 (SCH_3, α-**54**).

ESI-MS: 565.57 (100, $[M+H]^+$).

5.2.38 3',5'-Dimethylbenzoyl protected 2-thiomethylpyrimidine nucleoside (55)



51 (0.22 g, 0.85 mmol) in 2.2 ml dry pyridine was added dropwise to dimethylbenzoyl chloride (0.34 g, 1.79 mmol) at 0 °C under Ar. It was stirred for 16 h. EtOAc was added and it was washed once with 5 % HCl solution in H_2O and twice with H_2O, dried over Na_2SO_4 and evaporated under reduced pressure. After CC (SiO_2, hexane/EtOAc 2:1), **55** (0.34 g, 77 %) was obtained as a waxlike solid.

R_f (SiO_2, Hexan/EtOAc 2:1): 0.54

^{1}H-NMR (500 MHz, $CDCl_3$): δ 8.41 (*d*, *J* = 5.0 Hz, 1 H, HC(6), α-**55**), 8.37 (*d*, *J* = 5.0 Hz, 1 H, β-**55**), 7.73-7.65 (*m*, 4 H, arom. H_{meta}, α/β-**55**), 7.26-7.20 (*m*, 4 H, arom. H_{ortho}, α/β-**55**), 6.93 (*d*, *J* = 5.0 Hz, 1 H, HC(5), α/β-**55**), 5.52-5.50 (*m*, 1 H, HC(1'), α/β-**55**), 4.78-4.73 (*m*, 1 H, HC(3'), α-**55**), 4.71-4.63 (*m*, 1 H, HC(3'), β-**55**), 4.59-4.45 (*m*, 2 H, $H_2C(5')$, α/β-**55**), 4.38 (*td*, *J* = 4.3, 2.2 Hz, 1 H, HC(4'), α/β-**55**), 3.18 (*dd*, *J* = 13.8, 7.3 Hz, 1 H, HCC(4), α-**55**), 3.00 (*d*, *J* = 6.2 Hz, 2 H, $H_2CC(4)$, β-**55**), 2.70 (*dt*, *J* = 14.1, 7.1 Hz, 1 H, HC(4), α-**55**), 2.56 (*s*, 3 H, H_3CS, β-**55**),), 2.54 (*s*, 3 H, H_3CS, α-**55**), 2.40-2.39 *(m*, 1 H, HC(2'), α/β-**55**), 2.37 (*s*, 12 H, H_3C, β-**55**), 2.35 (*s*, 12 H, H_3C, α-**55**), 2.16-2.08 (*m*, 1 H, HC(2'), α/β-**55**).

^{13}C-NMR (125 MHz, $CDCl_3$): δ 172.5 (C(2), α-**55**), 172.4 (C(2), β-**55**), 167.2, 167.1 (CO, α/β-**55**), 166.6 (C(4), α-**55**), 166.5 (C(4), β-**55**), 157.0 (C(6), α-**55**), 156.9 (C(6), β-**55**), 138.2, 138.1 (arom. C_{ipso}, α/β-**55**), 135.0, 134.9 (arom. C_{meta}, α/β-**55**), 129.7, 129.5 (arom. C_{para}, α/β-**55**), 127.9, 127.4 (arom. C_{ortho}, α/β-**55**), 116.6 (C(5), β-**55**), 116.3 (C(5), α-**55**), 82.7 (C(4'), β-**55**), 81.6 (C(4'), α-**55**), 77.8 (C(3'), β-**55**), 77.7 (C(3'), α-**55**), 76.9 (C(1'), β-**55**), 76.6 (C(1'), α-**55**), 64.6 (C(5'), β-**55**), 64.5 (C(5'), α-**54**), 43.8 (*C*C(4), α-**55**), 43.2 (*C*C(4), β-**55**), 38.5 (C(2'), β-**55**), 37.5 (C(2'), α-**55**), 21.2, 21.1 (CH_3, α/β-**55**), 14.1 (SCH_3, α-**55**), 14.0 (SCH_3, β-**55**).

ESI-MS: 521.51 (100, $[M+H]^+$), 543.57 (15, $[M+Na]^+$).

5.2.39 CE PA dU

Cyanoethylphosphoramidite dU was synthesised in two steps from commercially available 2'-deoxyuridine in good yields, according to well established literature procedures.[111]

5.2.40 Oligonucleotide Synthesis

Oligonucleotides were synthesised on a DNA-synthesiser (*Expedite* 8900, *Applied Biosystems*) in 1.0 μmol scale using standard specifications and protocol for automated 3'→5' synthesis with following modifications: The phosphoramidite concentrations were reduced from 0.1 M to 0.06 M. Furthermore, the more potent 5-(ethylthio)-1*H*-tetrazole was used as activator, instead of 1*H*-tetrazole. In addition, the coupling times for the artificial nucleotide dD as well as the abasic site were elongated from 90 sec to 12 min. Oligonucleotide syntheses were performed staring from commercially available nucleosides, bound to *controlled pore glas* (1000 Å) from Biosearch. The corresponding commercially available cyanoethyl phosphoramidites were used for the insertion of dA, dG and dC. To release the synthesised oligonucleotide strands from solid support and to chip all nucleobase and phosphate protecting groups, they were treated with oversaturated aq. NH_3-solution at 50 °C for 20 h. Purification was achieved by RP-HPLC on a semi-preperative YMC basic column with a H_2O-MeCN gradient. The oligonucleotides were analysed by MALDI-TOF-MS (Matrix: 3-HPA). The oligonucleotide concentrations were determined by the *Lambert-Beer* law measuring the optical density at 260 nm.

6 References

[1] Voet, D.; Voet, J. G. *Biochemistry*, 3rd ed.; John Wiley & sons, New York, 2004.

[2] Miller, S. L.; Urey, H. C. *Science* **1959**, *130*, 245-251.

[3] Stetter, K. O. *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)*, Wiley, Chichester, U. K., 1996.

[4] Bada, J. L.; Lazcano, A. *Science* **2002**, *296*, 1982-1983.

[5] Willis, C.; Bada, J. *The Spark of Life: Darwin and the Primeval Soup*, Perseus, Cambridge, 2000.

[6] Bennett, J. O.; Shostak, G. S. *Life in the Universe*, 2nd ed.; Pearson: Addison-Wesley, 2007.

[7] Knoll, A. H.; Barghoorn, E. S. *Science* **1977**, *198*, 596.

[8] Miller, S. L. *Science* **1953**, *117*, 528-529.

[9] Gilbert, W. *Nature* **1986**, *319*, 618.

[10] Altman, S. *Cell* **1984**, *36*, 237-239.

[11] Guerrier-Takada, C.; Altman, S. *Science* **1984**, *223*, 285-286.

[12] Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147-157.

[13] Zaug, A. J.; Cech, T. R. *Science* **1986**, *231*, 470-475.

[14] White III, H. B. *J. Mol. Evol*. **1976**, *7*, 101-104.

[15] Orgel, L. E. *J. Mol. Biol*. **1968**, *38*, 381-393.

[16] Sharp, P. A. *Cell* **1985**, *42*, 397-400.

[17] Oró, J. *Biochem. Biophys. Res. Commun*. **1960**, *2*, 407-412.

[18] Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *Science* **1966**, *154*, 784-785.

[19] (a) Ferris, J. P.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1968**, *33*, 693-704. (b) Ferris, J. P.; Zamek, O. S.; Altbuch, A. M.; Freiman, H. *J. Mol. Evol.* **1974**, *3*, 301-309.

[20] Robertson, M. P.; Miller, S. L. *Nature* **1995**, *375*, 772-774.

[21] Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1972**, *67*, 25-33.

[22] Orgel, L. E. *Crit. Rec. Biochem. Mol. Biol.* **2004**, *39*, 99-123.

[23] Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459*, 239-242.

[24] (a) Thaddeus, P. *Phil. Trans. R. Soc*. **2006**, *361*, 1681-1687. (b) Bryant, D. E.; Kee, T. P. *Chem. Commun.* **2006**, 2344-2346.

[25] Larralde, R.; Robertson, M. P.; Miller, S. L. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8185-8160.

[26] (a) Eschenmoser, A.; Dobler, M. *Helv. Chim. Acta* **1992***, 75*, 218-259. (b) Eschenmoser, A.; Loewenthal, E. *Chem. Soc. Rev*. **1992**, *21*, 1-16. (c) Eschenmoser, A. *Origins Life Evol. Biosphere* **2004**, *34*, 277-306.

[27] (a) Böhringer, M.; Roth, H.-J.; Hunziker, J.; Göbel, M.; Krishnan, R.; Giger, A.; Schweizer, B.; Schreiber, J.; Leumann, C.; Eschenmoser, A. *Helv. Chim. Acta* **1992**, *75*, 1416-1477. (b) Egli, M.; Pallan, P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A. *J. Am. Chem. Soc.* **2006**, *128*, 10847-56.

[28] Crey-Desbiolles, C.; Ahn. D.-R.; Leumann, C. J. *Nuclic Acids Res*. **2005**, *33*, e77/1-7.

[29] Schöning, K.-U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *209*, 1347-1351.

[30] Ueda, N.; Kawabata, T.; Takemoto, K. *J. Hererocycl. Chem*. **1971**, *8*, 827-829.

[31] Schwartz, A. W.; Orgel, L. E. *Science* **1985**, *228*, 585-587.

[32] Nelson, K. E.; Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3868-3871.

[33] Westheimer, F. H. *Science* **1987**, *235*, 1173-1178.

[34] Usher, D. A. *Nature New Biol.* **1972**, *235*, 207-208.

[35] Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33-37.

[36] Kolb, V. M.; Dworkin, J. P.; Miller, S. L. *J. Mol. Evol.* **1994**, *38*, 549-557.

[37] Wächtershäuser, G. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1134-1135.

[38] Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859-862.

[39] Siegel, J. S.; Tor, Y. *Org. Biomol. Chem.* **2005**, *3*, 1591-1592.

[40] Haeckel, E. *Generelle Morphologie der Organismen*, Reimer, Berlin, 1866.

[41] Miescher, F. *medizinisch-chemische Untersuchungen* **1871**, *4*, 441-460.

[42] Dahm, R. *Developmental Biology* **2005**, *278*, 274-288.

[43] Boveri, T. *Jena. Zeit. Naturwiss*. **1888**, *22*, 685-882.

[44] Johannsen, W. *Elemente der exakten Erblichkeitslehre*, Gustav Fischer, Jena, 1909.

[45] Griffith, F. *J. Hyg*. **1928**, *27*, 113.

[46] (a) Levene, P. A. *Biochem. Z*. **1909**, *17*, 121. (b) Levene, P. A.; Jacobs, W. A. *Ber. Chem. Ges*. **1909**, *xlii*, 2475 and 2703.

[47] Kay, L. E. *The Molecular Vision of Life*, Oxford University Press, 1992.

[48] Avery, O. T.; LacLeod, C. M.; McCarty, M. *J. Exp. Med***. 1944**, *79*, 137-159.

[49] (a) Vischer, E.; Zamenhof, S.; Chargaff, E. *J. Biol. Chem.* **1949**, *177*, 429. (b) Chargaff, E.; Zamenhof, S.; Brawerman, G.; Kerin, L. *J. Am. Chem. Soc*. **1950**, *72*, 3825.

[50] Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171*, 740-741.

[51] Wilkins, M. H.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *171*, 738-740.

[52] Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

[53] Bartlett, J. M. S.; Stirling, D. *PCR Protocols* **2003**, *226*, 3-6.

[54] *Science* **2001**, *291*, 1218.

[55] Saenger, W. *Principles of Nucleic Acids Structure*, Springer, Berlin, 1984.

[56] Pauling, L. *The Nature of the Chemical Bond*, Cornell Univ. Press, Ithaca, N. Y., 1978.

[57] Elguero, J.; Marzin, C.; Katritzky, A. R.; Linda, P. *The Tautomerism of Heterocycles, Academic Press*, New York, 1976.

[58] Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J. *J. Org. Chem.* **1996**, *61*, 266-286.

[59] Velikyan, I.; Acharya, S.; Trifonova, A.; Foeldesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2001**, *123*, 2893-2894.

[60] Lindahl, T. *Nature* **1993**, *362*, 709-715.

[61] Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7933-7938.

[62] Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, *29*, 2532-2537.

[63] Lindahl, T.; Karlström, O. *Biochemistry* **1973**, *12*, 5151-5154.

[64] Lindahl, T. *J. Biol. Chem*. **1967**, *242*, 1970-1973.

[65] (a) Klein, W.; Thannhauser, S. J. *Z. physiol. Chem*. **1933**, *218*, 173-180. (b) Klein, W.; Thannhauser, S. J. *Z. physiol. Chem*. **1934**, *224*, 252-260.

[66] Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *172*, 156-157.

[67] Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680-686.

[68] Bugg, C. E.; Thomas, J.M.; Sundaralingam, M.; Rao, S. T. *Biopolymers* **1971**, *10*, 175-219.

[69] Hanlon, S. *Biochem. Biophys. Res. Common*. **1966**, *23*, 861-867.

[70] Herskovits, T. T. *Arch. Biochem. Biophys*. **1962**, *97*, 474-484.

[71] Moran, S.; Ren, R. X.-F.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056-2057.

[72] Crick, F. H. C. *J. Mol. Biol*. **1966**, *19*, 548-555.

[73] Varani, G.; McClain, W. *EMBO Rep*. **2000**, *1*, 18-23.

[74] Hoogsteen, K. *Acta Cryst*. **1963**, *16*, 907-916.

[75] Johnson, J. E.; Smith, J. S.; Kozak, M. L.; Johnson, F. B. *Biochimie* **2008**, *90*, 1250-1263.

[76] Graziewicz, M. A.; Zastawny, T. H.; Olinski, R.; Tudek, B. *Mutat. Res.-DNA Repair*. **1999**, *434*, 41-52.

[77] Greenberg, M. M.; Hantosi, Z.; Wiederholt, C. J.; Rithner, C. D. *Biochemistry* **2001**, *40*, 15856-15861.

[78] Moss, R. J.; Petrie, C. R.; Meyer, R. B.; Nord, L. D.; Willis, R. C.; Smith, R. A.; Larson, S. B.; Kini, G. D.; Robins, R. K. *J. Med. Chem*. **1988**, *31*, 786-790.

[79] Fortin, H.; Tomasi, S.; Delcros, J.-G.; Bansard, J.-Y.; Boustie, J. *ChemMedChem* **2006**, *1*, 189-196.

[80] Hysell, M.; Siegel, J. S.; Tor, Y. *Org. Biomol. Chem.* **2005**, *3*, 2946-2952.

[81] Haraguchi, K.; Delaney, M. O.; Wiederholt, C. J.; Sambandam, A.; Hantosi, Z.; Greenberg, M. M. *J. Am. Chem. Soc.* **2002**, *124*, 3263-3269.

[82] Delanay, M. O.; Wiederholt, C. J; Greenberg, M. M.; *Angew. Chem. Int. Ed.* **2002**, *41*, 771-773.

[83] Gensler, W. J.; Chan, S.; Ball, D. B. *J. Am. Chem. Soc*. **1975**, *97*, 436-437.

[84] Ober, M.; Linne, U.; Gierlich, J.; Carell, T. *Angew. Chem. Int. Ed.* **2003**, *42*, 4947-4951.

[85] Michelson, A. M.; Todd, A. R. *J. Chem. Soc*. **1955**, 2632-2638.

[86] (a) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc*. **1965**, *87*, 3526-3527. (b) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc.* **1966**, *88*, 5319-5324.

[87] Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859-1862.

[88] Ladbury, J. E.; Chowdhry, B. Z. *Biocalorimetry- Applications of Calorimetry in the Biological Sciences*, John Wiley & Sons, New York, 1998.

[89] Feig, A. L. *Biopolymers* **2007**, *87*, 293-301.

[90] Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*, University Science Books, California, 2006.

[91] Bruylants, G.; Boccongelli, M.; Snoussi, K.; Bartik, K. *Biochemistry* **2009**, *48*, 8473-8482.

[92] (a) Lang, B. E.; Schwarz, F. P. *Biophys. Chem*. **2007**, *131*, 96-104. (b) Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T. Jr. *Biochemistry* **1999**, *38*, 8409-8422. (c) Takach, J. C.; Mikulecky, P. J.; Feig, A. L. *J. Am. Chem. Soc*. **2004**, *126*, 6530-6531. (d) Mikulecky, P. J.; Feig, A. L. *Biochemistry* **2006**, *45*, 604-616.

[93] Haraguchi, K.; Greenberg, M. M. *J. Am. Chem. Soc*. **2001**, *123*, 8636-8637.

[94] Robertson, M. P.; Levy, M.; Miller, S. L. *J. Mol. Evol.* **1996**, *43*, 543–550.

[95] Shapiro, R. *Proc. Natl. Acad. Sci.* **1999**, *96*, 4396-4401.

[96] Berstis, L. M.Sc. Thesis, University of Zurich, 2009.

[97] Löpfe, M. Ph.D. Thesis, University of Zurich, 2009.

[98] Bischof, B. Ph.D. Thesis, University of Zurich, 2009.

[99] Delaney, M. O.; Greenberg, M. M. *Chem. Res. Toxicol.* **2002**, *15*, 1460-1465.

[100] Cupps, T. L.; Wise, D. S.; Townsend, L. B. *J. Org. Chem.* **1986**, *51*, 1058-1064.

[101] a) Wittig, G.; Schöllkopf, U. *Chem. Berichte* **1954**, *87*, 1318. b) Wittig, G.; Haag, W. *Chem. Berichte* **1954**, *88*, 1654-1666.

[102] Strekowski, L.; Wydra, R. L.; Janda, L.; Harden, D. B. *J. Org. Chem.* **1991**, *56*, 5610-5614.

[103] Steinauer, A. Master Thesis, University of Zurich, 2012.

[104] McKenna, J. M.; Halley, F.; Souness, J. E.; McLay, I. M.; Pickett, S. D.; Collis, A. J.; Page, K.; Ahmed, I. *J. Med. Chem.* **2002**, *45*, 2173-84.

[105] Song, J.; Hollingsworth, R. I. *Tetrahedron: Asymmetry* **2001**, *12*, 387-391.

[106] Walker, J. a.; Chen, J. J.; Wise, D. S.; Townsend, L. B. *J. Org. Chem.* **1996**, *61*, 2219-2221.

[107] Herrera, A.; Martinez-Alvarez, R.; Ramiro, P.; Molero, D.; Almy, J. *J. Org. Chem.* **2006**, *71*, 3026-3032.

[108] Ahaidar, A.; Fernández, D.; Danelón, G.; Cuevas, C.; Manzanares, I.; Albericio, F.; Joule, J. A.; Álvarez, M. *J. Org. Chem.* **2003**, *68*, 10020-10029.

[109] Gait, M. J. *Oligonucleotide synthesis – a practical approach*, IRL Press, Oxford, 1984.

[110] Greenberg, M. M.; Hwang, J.-T. *J. Am. Chem. Soc*. **1999**, *121*, 4311-4315.

[111] Ren, R. X.-F.; Chaudhuri, N. C.; Paris, P. L.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 7671-7678.

[112] Matsukawa, T.; Ohta, B. *Yakugaku Zasshi* **1949**, *69*, 491-493.

[113] Greene, T. W.; Wuts, P. G. M. *Protective Groups in Organic Synthesis*, 2nd ed.; John Wiley & Sons, New York, 1991.

[114] Herrera, A.; Martinez-Alvarez, R.; Chioua M.; Chatt, R.; Chioua, R.; Sanchez, A.; Almy, J. *Tetrahedron* **2006**, 62, 2799-2811.

[115] McOmie, J. F. W.; West, D. E. *Org. Synth. Coll.* **1973**, *5*, 412.

[116] Jung, M. E.; Lyster, M. A.; *J. Org. Chem*. **1977**, *42*, 3761-3764.

[117] Reese, C. B. in *Protective Groups in Organic Chemistry*, McOmie, J. F. W., Ed., Plenum Press, London, 1973.

[118] Bhat, C.C. in *Synthetic Procedures in Nucleic Acid Chemistry*, Zorbach, W.W.; Tipson, R.S., Eds., John Wiley & Sons, New York, 1968.

[119] Rolland, V.; Kotera, M.; Lhomme, J. *Synth. Comm.* **1997**, *27*, 3505-3511.

[120] Hoffer. M. *Chem. Ber.* **1960**, *93*, 2777.

[121] Grünefeld, P.; Richert, C. *J. Org. Chem.* **2004**, *69*, 7543-7551.

[122] Shigdel, U. K.; He, C. *J. Am. Chem. Soc.* **2008**, *130*, 17634-17634.

[123] Tejima, S.; Maki, T.; Akagi, M. *Chem. Pharm. Bull.* **1964**, *12*, 528-532.

[124] Osornio, Y. M.; Cruz-Almanza, R.; Jiménez-Montaño, V.; Miranda, L. D. *Chem. Commun.* **2003**, 2316 - 2317.

[125] Zard, S. Z. in *Radical in Organic Synthesis*, Renaud, P.; Sibi, M., Eds., Wiley VCH, Weinheim, 2001.

[126] Huang, H.; Greenberg, M. M. *J. Org. Chem.* **2008**, *73*, 2695-2703.

[127] Takeshita, M.; Chang, C. N.; Johnson, F.; Will, S.; Grollman, A. P. *J. Biol. Chem.* **1987**, *262*, 10171-10179.

[128] Ikemoto, N.; Lo, L.-C.; Kim, K. O.; Berova, N.; Nakanishi, K. *Carbohydr. Res.* **1993**, *239*, 11-33.

[129] Tribolet, R.; Sigel, H. *Eur. J. Biochem.* **1987**, *163*, 353-363.

[130] Martin, R. B. *Science* **1963**, *139*, 1198-1199.

[131] Vesnaver, G.; Breslauer, K. J. *Proc. Natl. Acad. Sci.* **1991**, *88*, 3569-3573.

[132] Navidi, W. *Statistics for Engineers and Scientists*, 2nd ed.; McGraw Hill, 2008.

[133] Kool, E. T. *Chem. Rev*. **1997**, *97*, 1473-1487.

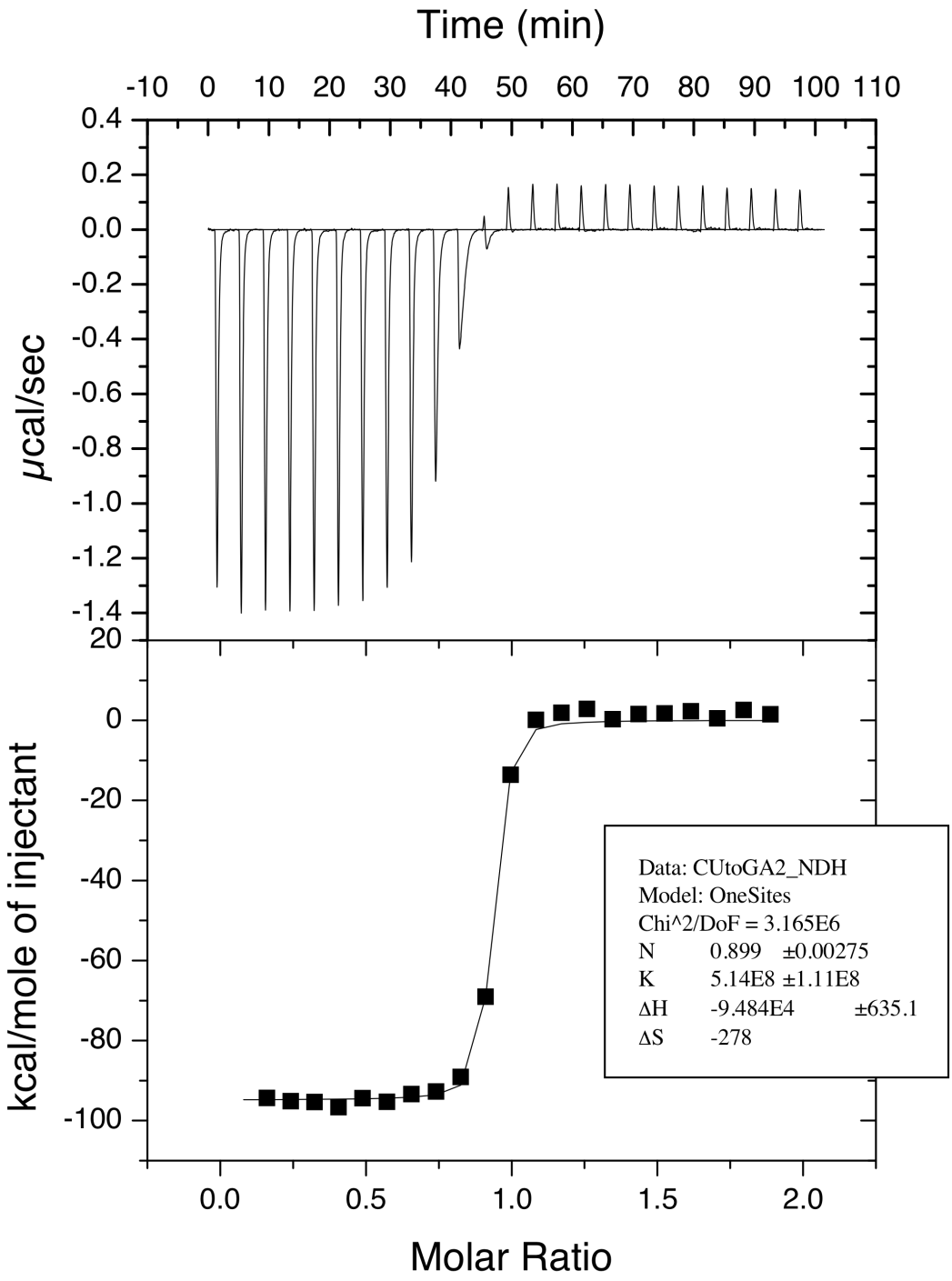
[134] Guckian, K.; Schweitzer, B. A.; Ren, X.-F.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc*. **1996**, *118*, 8182-8183.

[135] Bredereck, H.; Effenberger, F.; Simchen, G. *Chem. Ber*. **1962**, *135*, 1078-1080.

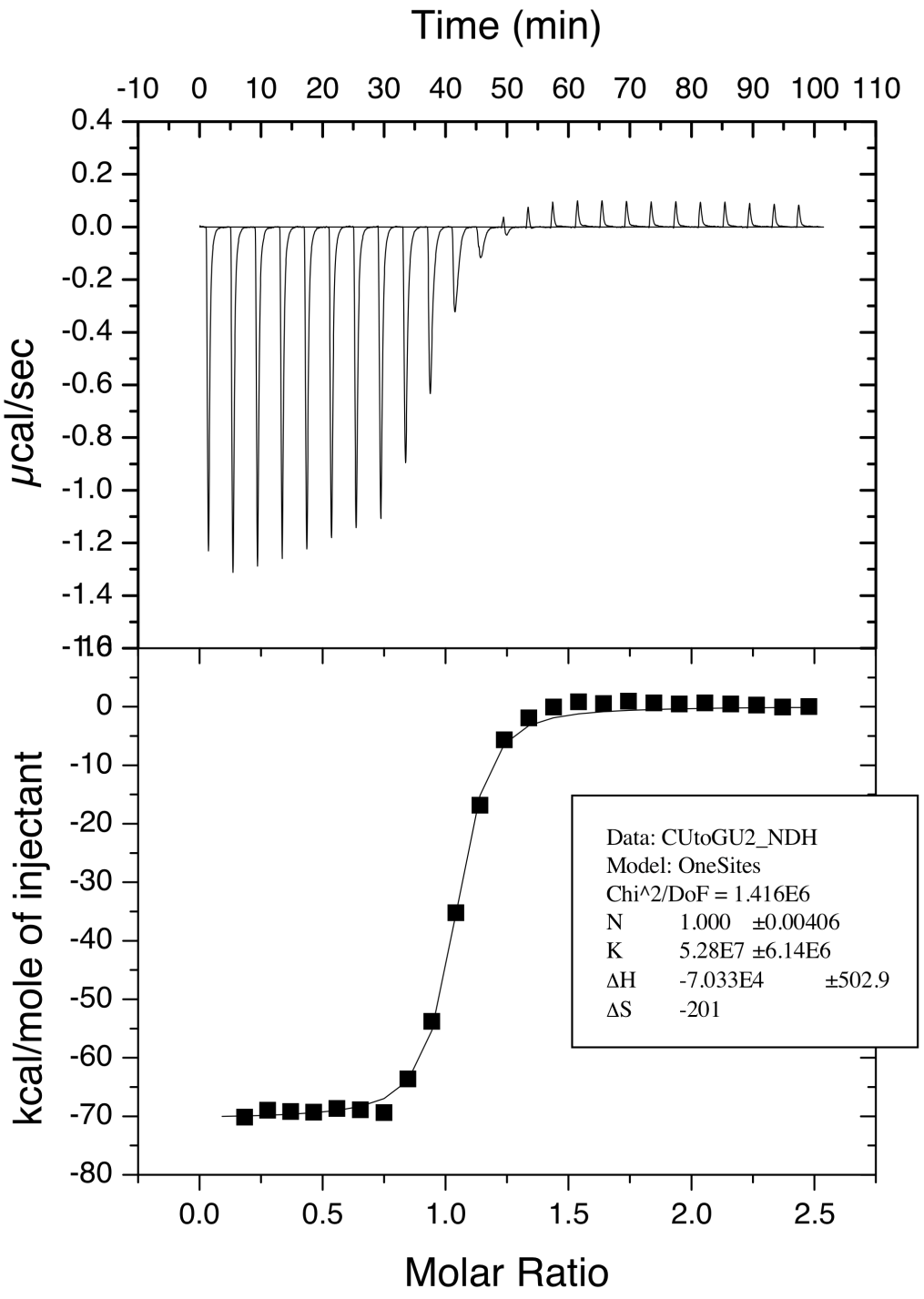
Appendix

ITC Data

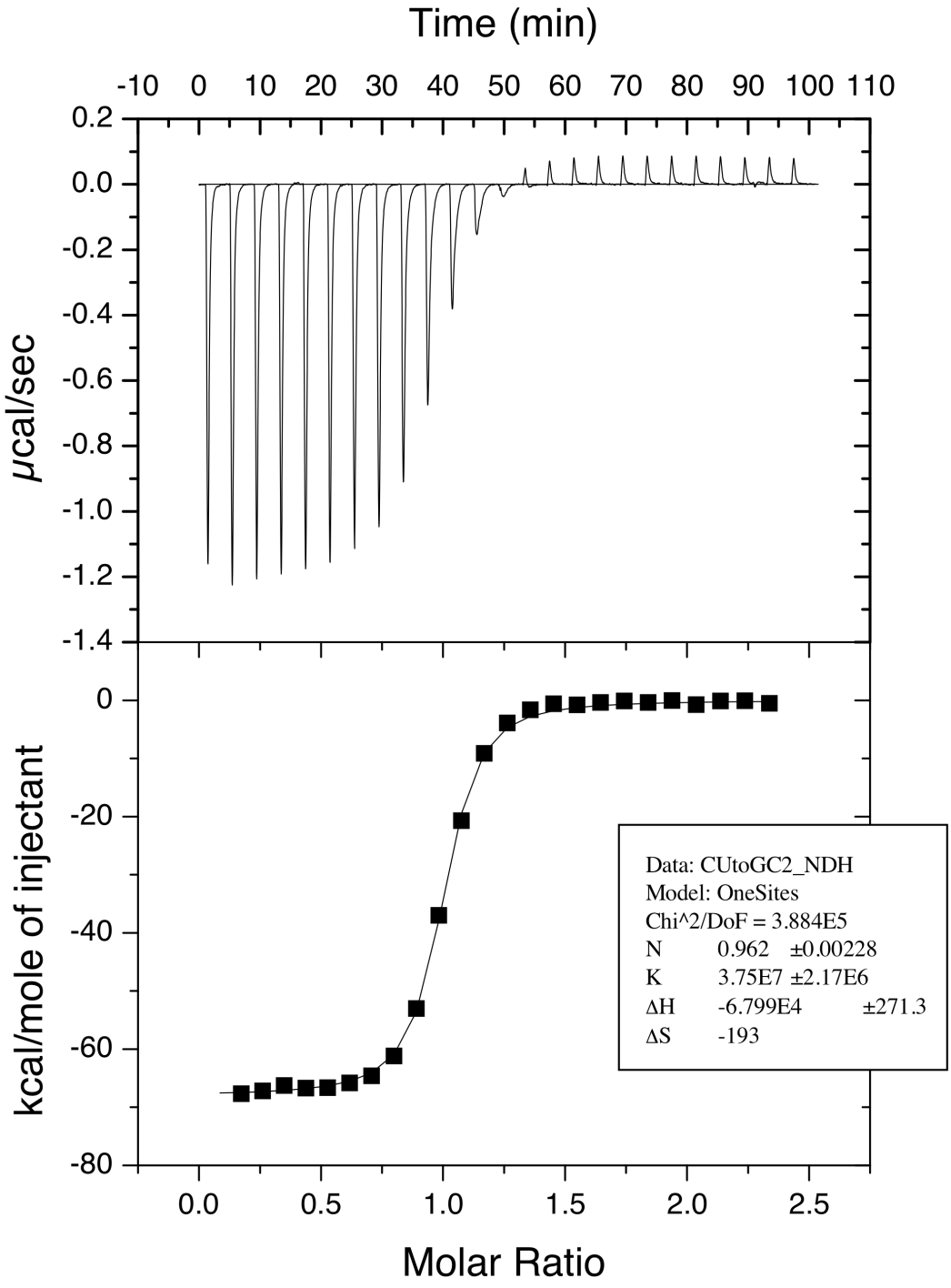
C_U to G_A



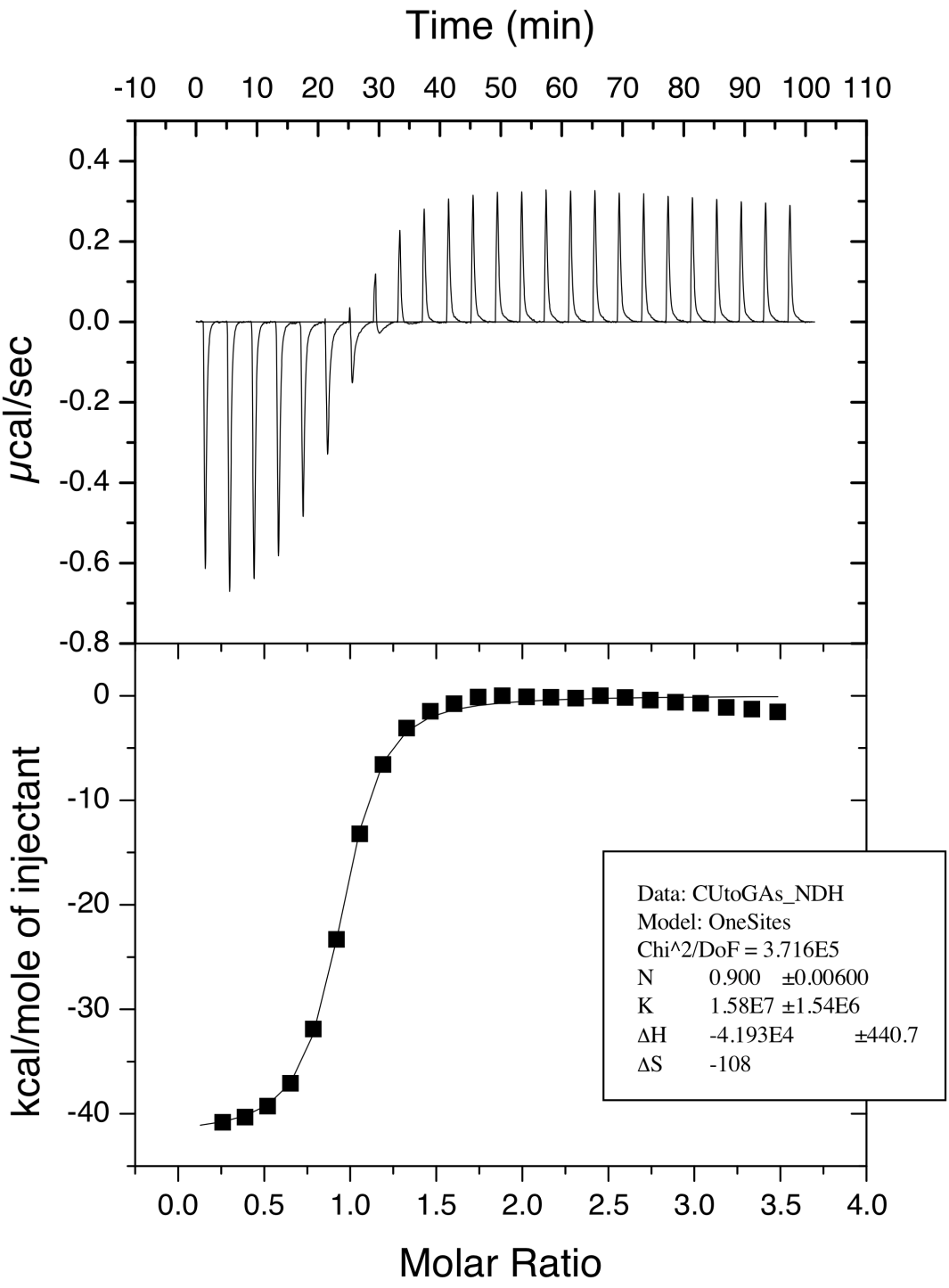
C_U to G_U


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
µcal/sec
0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4 -1.6
kcal/mole of injectant
0 -10 -20 -30 -40 -50 -60 -70 -80
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CUtoGU2_NDH
Model: OneSites
Chi^2/DoF = 1.416E6
N 1.000 ±0.00406
K 5.28E7 ±6.14E6
ΔH -7.033E4 ±502.9
ΔS -201

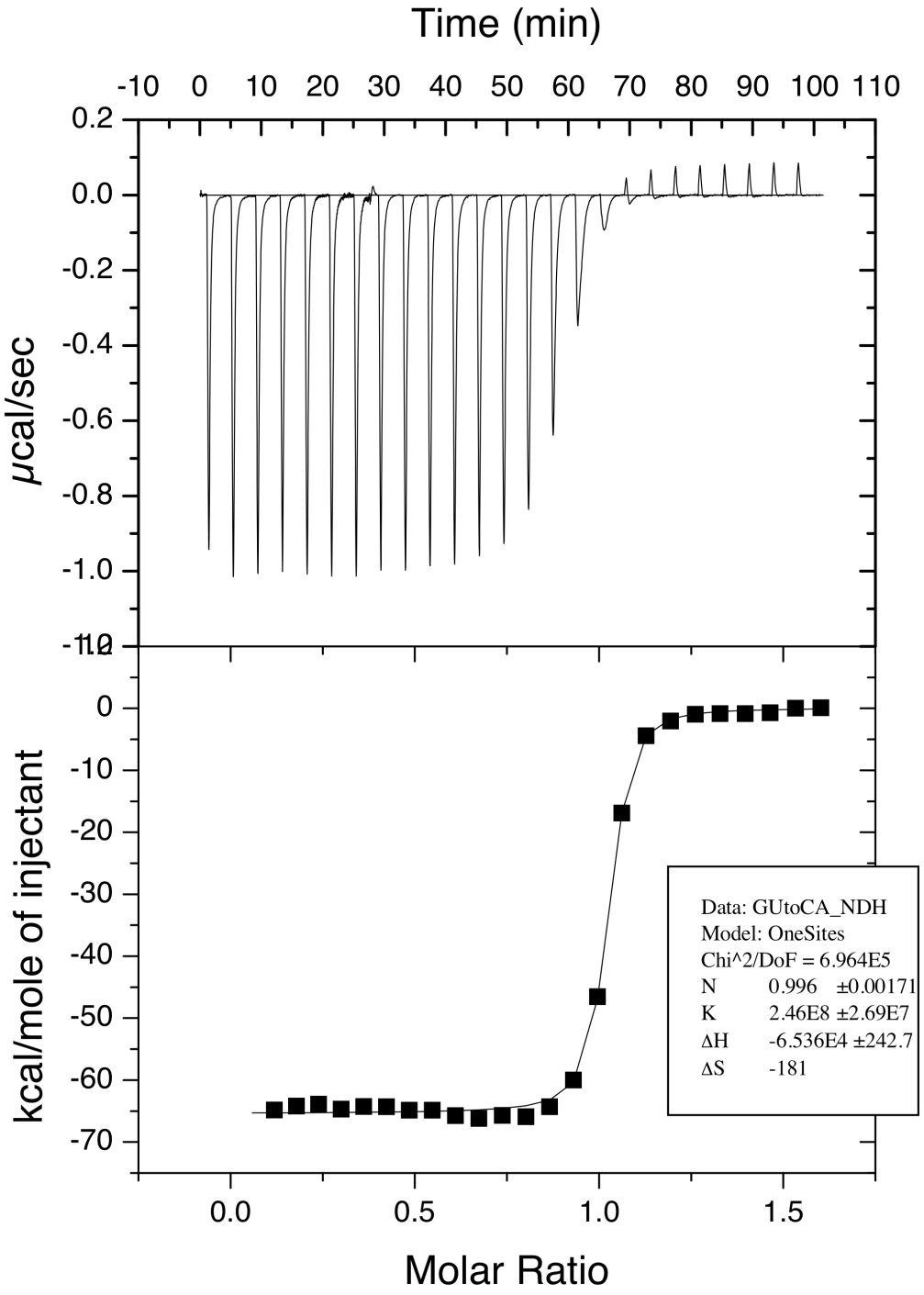
C_U to G_C


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
µcal/sec
0
-20
-40
-60
-80
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CUtoGC2_NDH
Model: OneSites
Chi^2/DoF = 3.884E5
N 0.962 ±0.00228
K 3.75E7 ±2.17E6
ΔH -6.799E4 ±271.3
ΔS -193

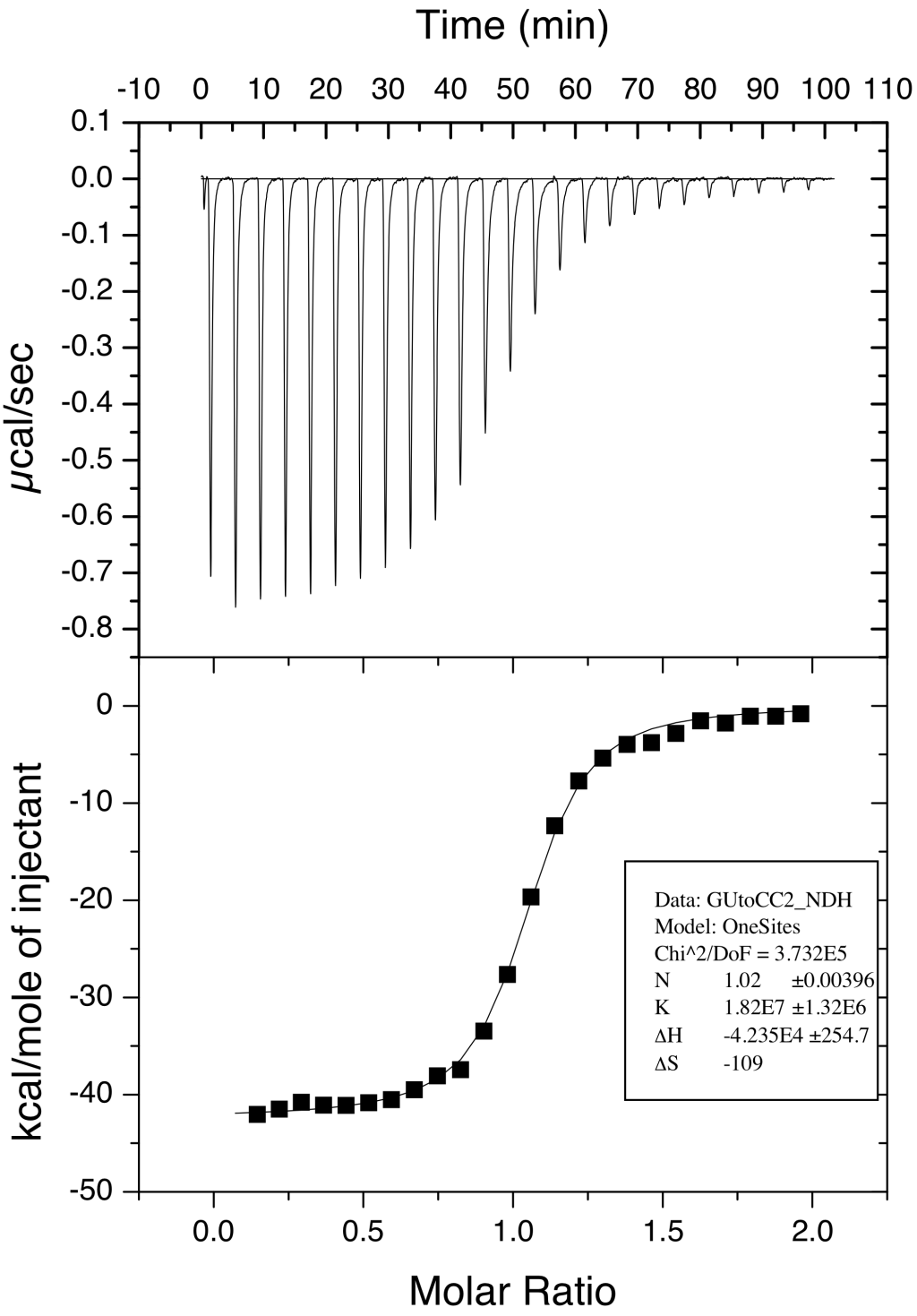
C_U to G_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Molar Ratio
Data: CUtoGAs_NDH
Model: OneSites
Chi^2/DoF = 3.716E5
N 0.900 ±0.00600
K 1.58E7 ±1.54E6
ΔH -4.193E4 ±440.7
ΔS -108

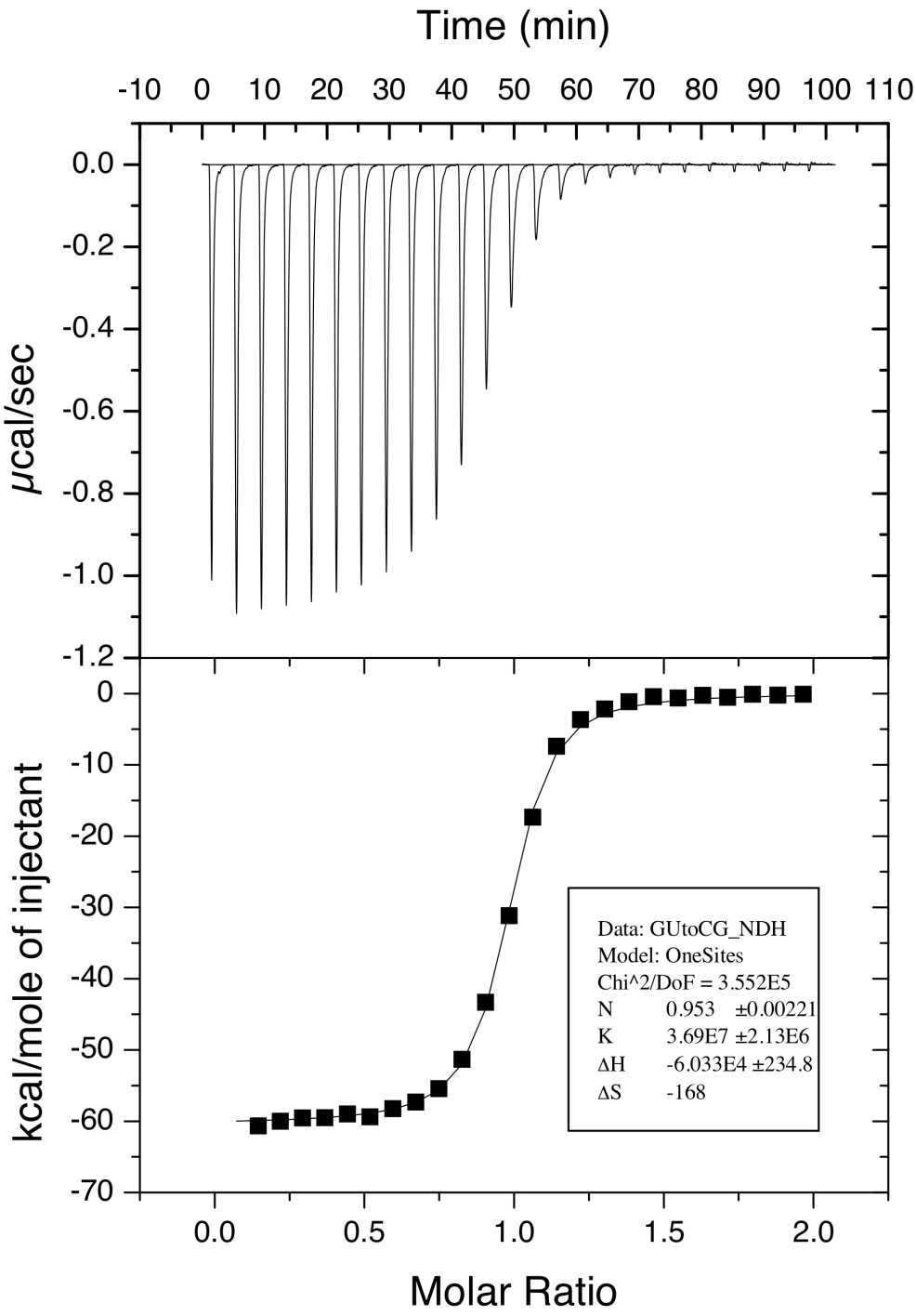
G_U to C_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
0
-10
-20
-30
-40
-50
-60
-70
kcal/mole of injectant
0.0
0.5
1.0
1.5
Molar Ratio
Data: GUtoCA_NDH
Model: OneSites
Chi^2/DoF = 6.964E5
N 0.996 ±0.00171
K 2.46E8 ±2.69E7
ΔH -6.536E4 ±242.7
ΔS -181

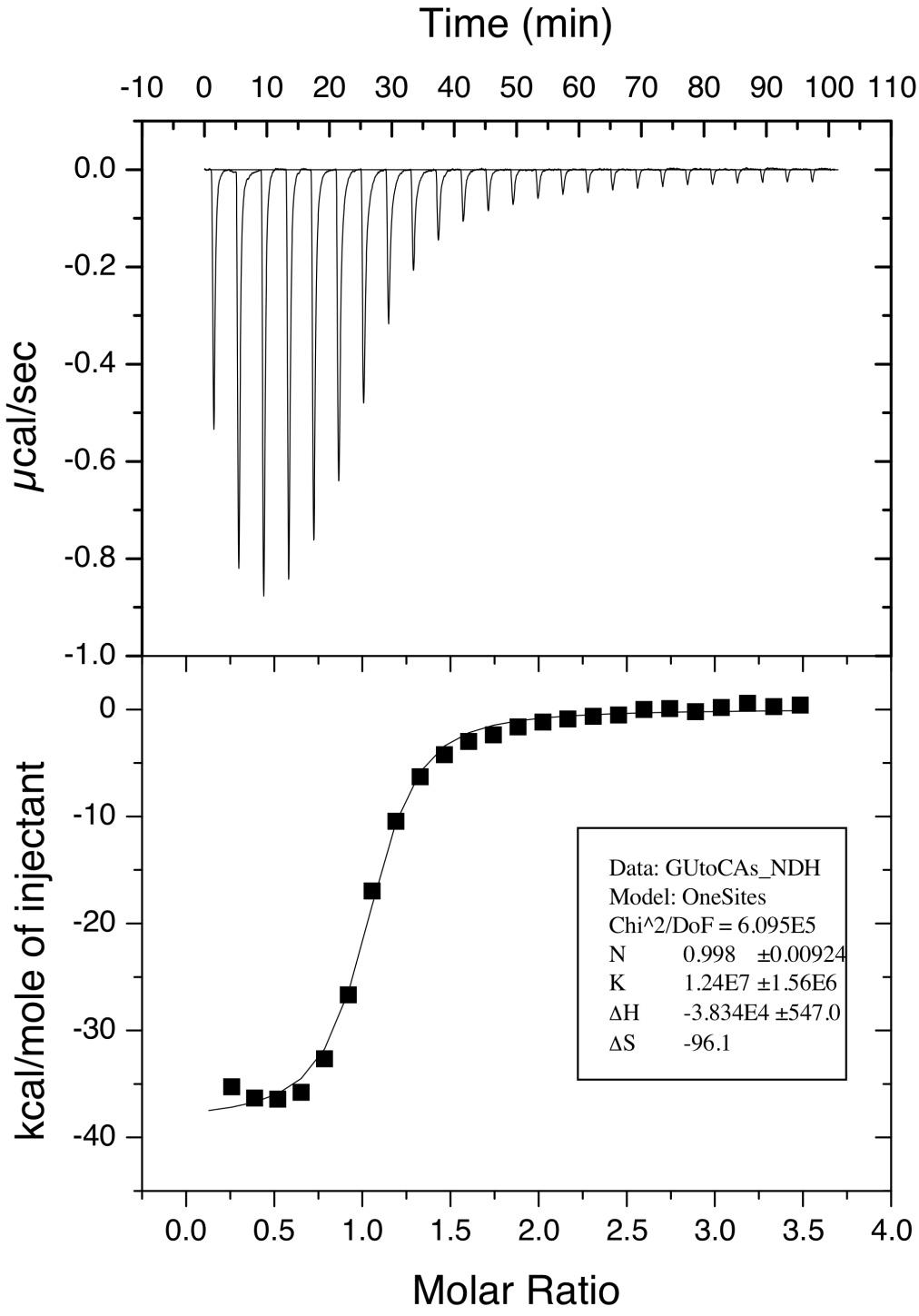
G_U to C_C


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GUtoCC2_NDH
Model: OneSites
Chi^2/DoF = 3.732E5
N 1.02 ±0.00396
K 1.82E7 ±1.32E6
ΔH -4.235E4 ±254.7
ΔS -109

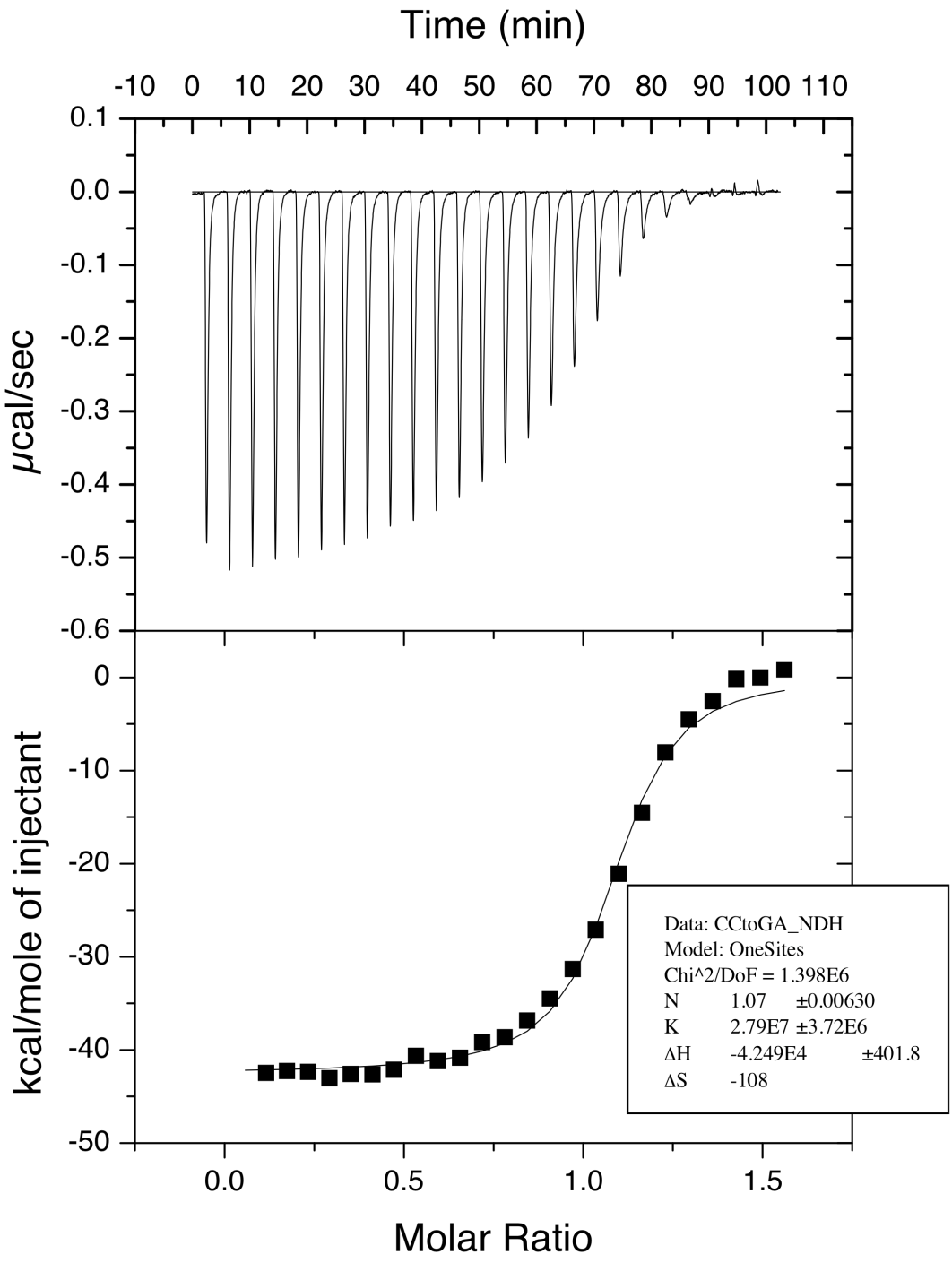
G_U to C_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
µcal/sec
0
-10
-20
-30
-40
-50
-60
-70
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GUtoCG_NDH
Model: OneSites
Chi^2/DoF = 3.552E5
N 0.953 ±0.00221
K 3.69E7 ±2.13E6
ΔH -6.033E4 ±234.8
ΔS -168

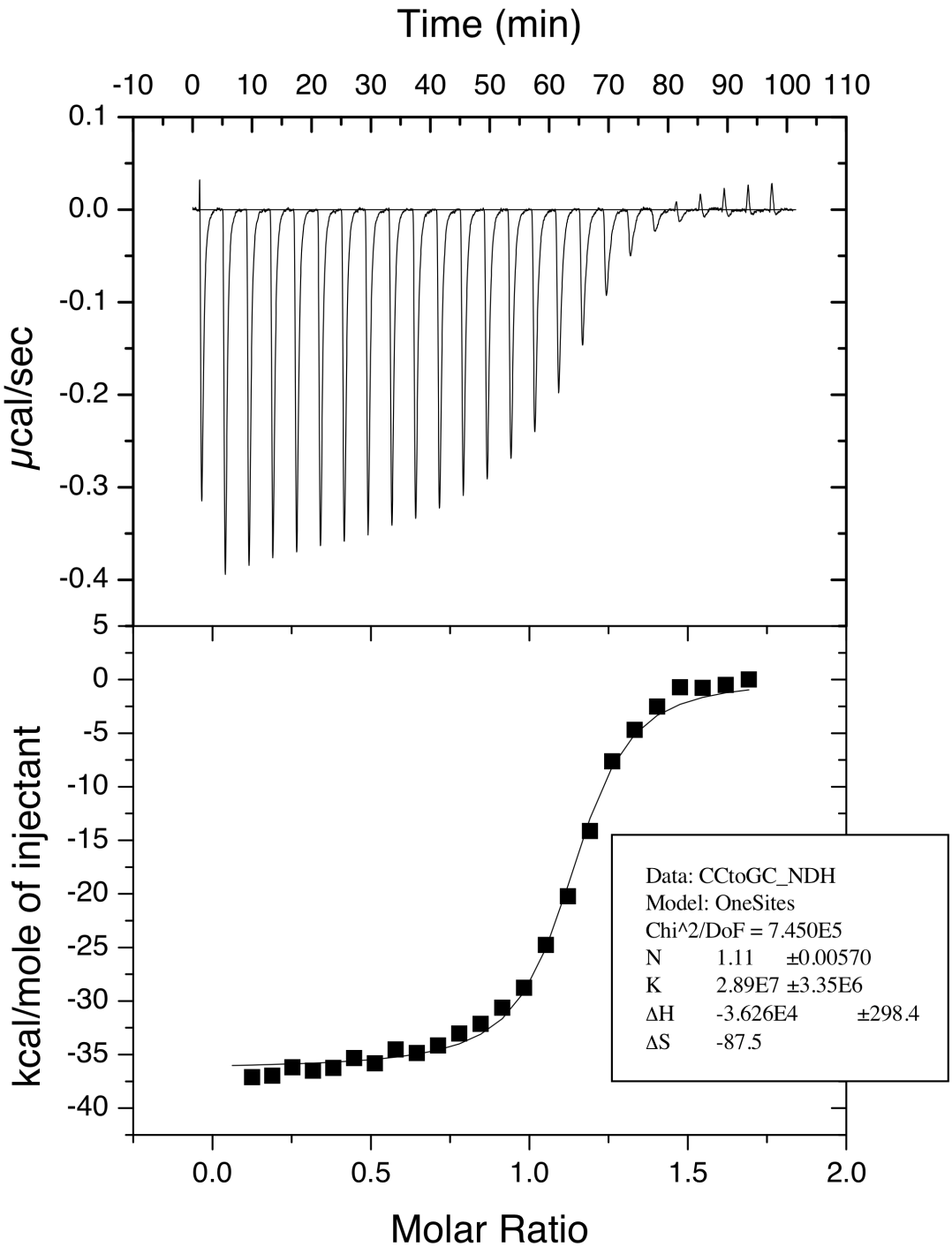
G_U to C_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
0
-10
-20
-30
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Molar Ratio
Data: GUtoCAs_NDH
Model: OneSites
Chi^2/DoF = 6.095E5
N 0.998 ±0.00924
K 1.24E7 ±1.56E6
ΔH -3.834E4 ±547.0
ΔS -96.1

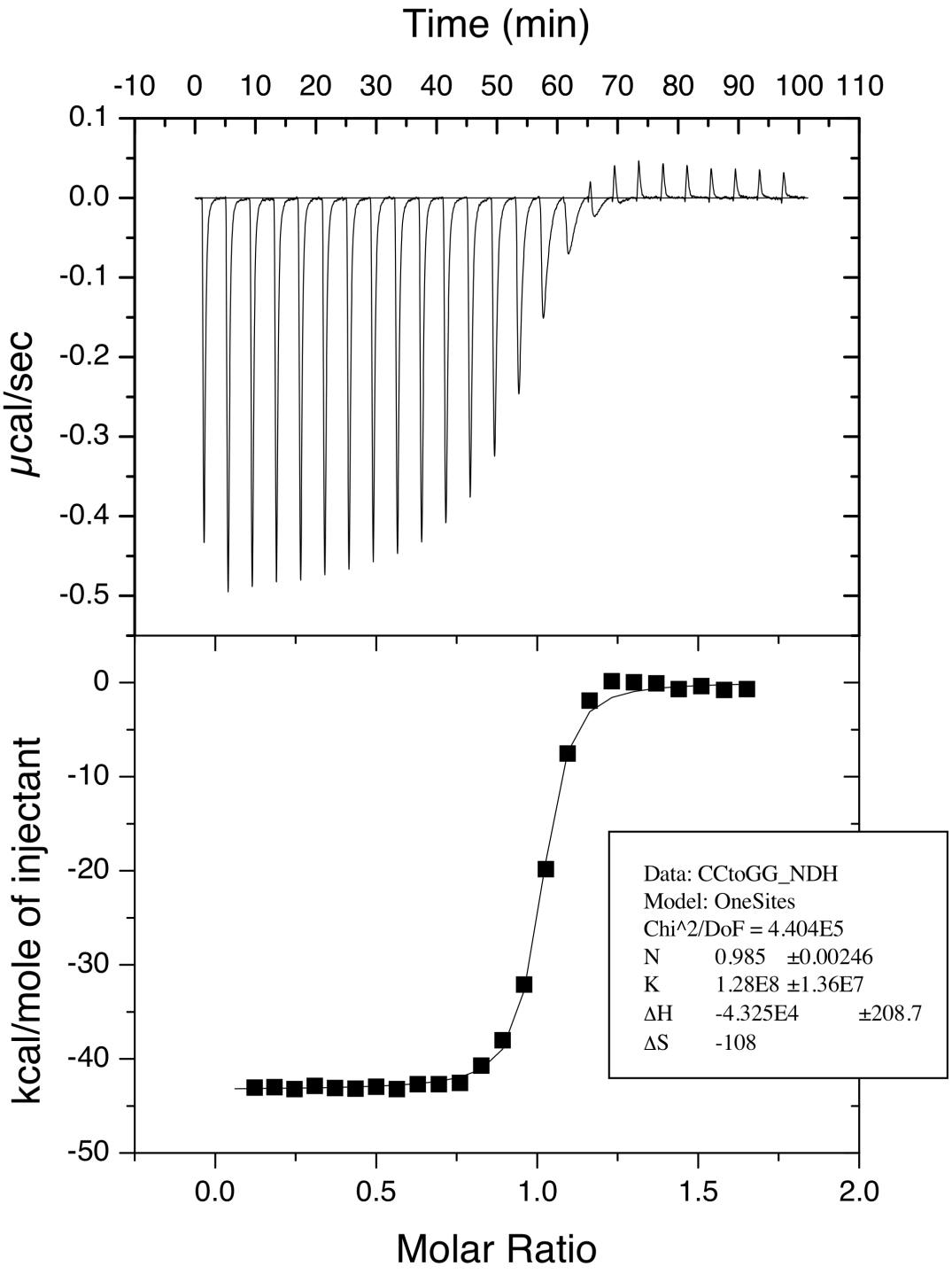
C_C to G_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0
0.5
1.0
1.5
Molar Ratio
Data: CCtoGA_NDH
Model: OneSites
Chi^2/DoF = 1.398E6
N 1.07 ±0.00630
K 2.79E7 ±3.72E6
ΔH -4.249E4 ±401.8
ΔS -108

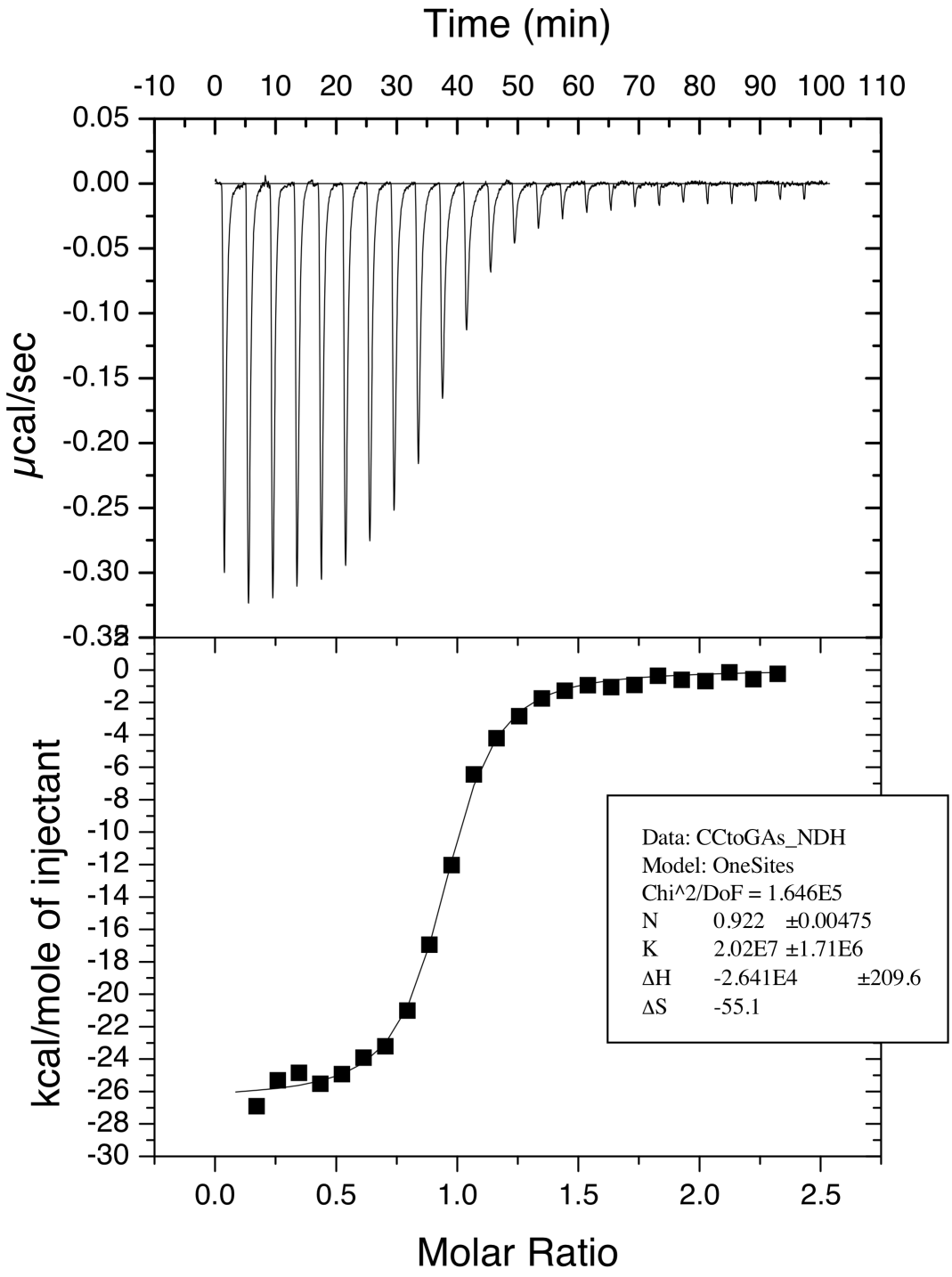
C_C to G_C


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
µcal/sec
5
0
-5
-10
-15
-20
-25
-30
-35
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CCtoGC_NDH
Model: OneSites
Chi^2/DoF = 7.450E5
N 1.11 ±0.00570
K 2.89E7 ±3.35E6
ΔH -3.626E4 ±298.4
ΔS -87.5

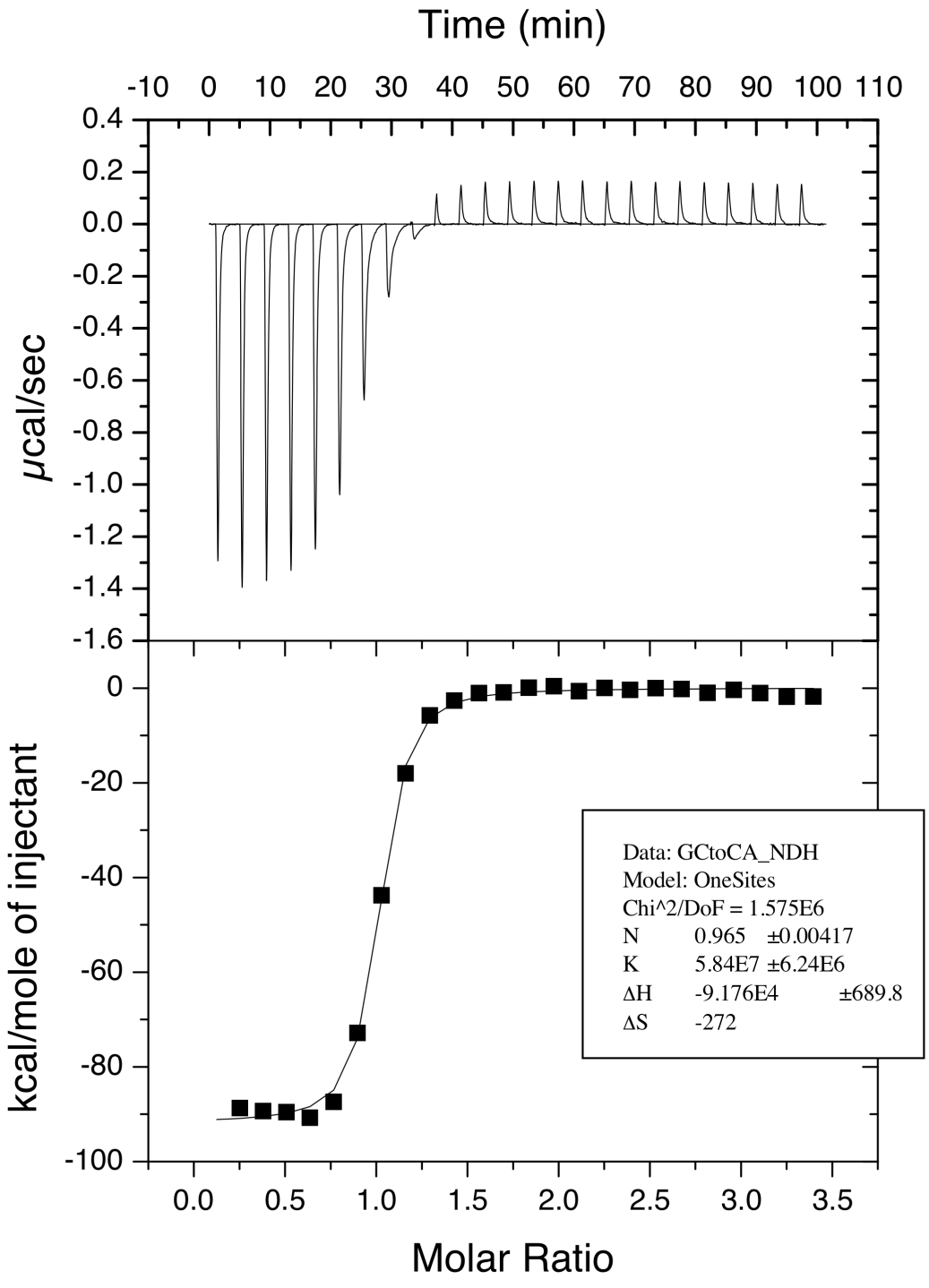
C_C to G_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CCtoGG_NDH
Model: OneSites
Chi^2/DoF = 4.404E5
N 0.985 ±0.00246
K 1.28E8 ±1.36E7
ΔH -4.325E4 ±208.7
ΔS -108

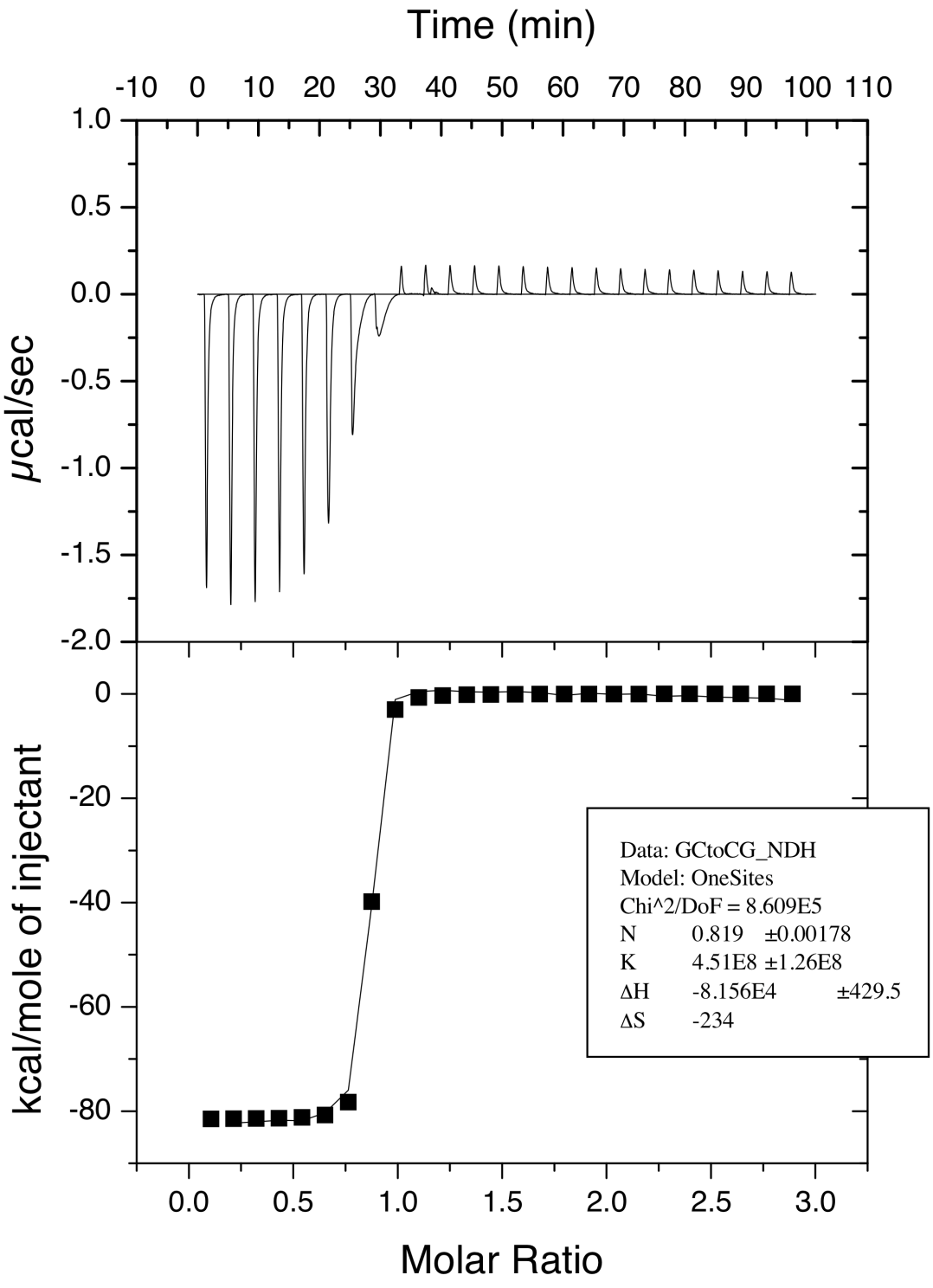
C_C to G_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
µcal/sec
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CCtoGAs_NDH
Model: OneSites
Chi^2/DoF = 1.646E5
N 0.922 ±0.00475
K 2.02E7 ±1.71E6
ΔH -2.641E4 ±209.6
ΔS -55.1

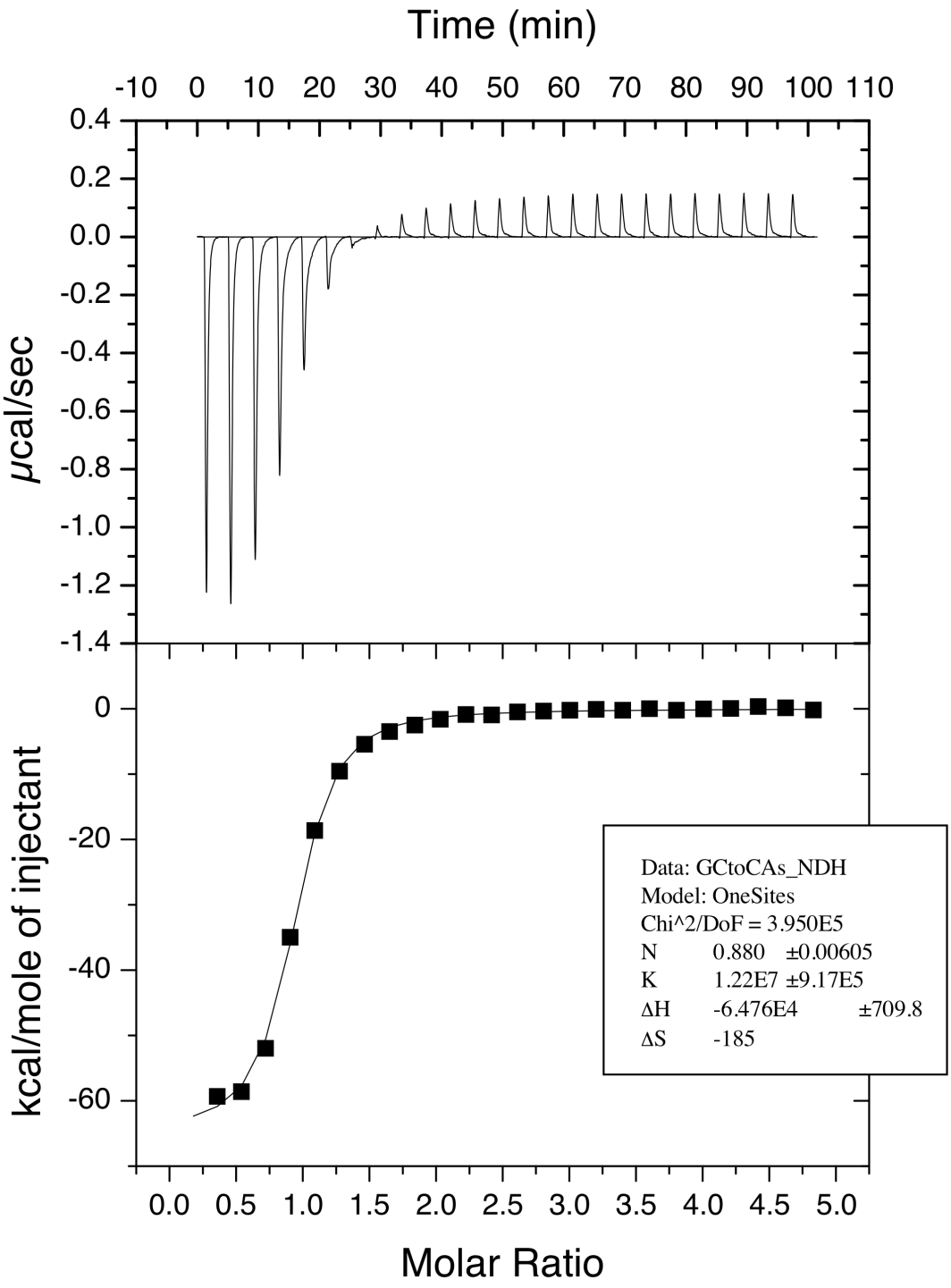
G_C to C_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
µcal/sec
0
-20
-40
-60
-80
-100
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Molar Ratio
Data: GCtoCA_NDH
Model: OneSites
Chi^2/DoF = 1.575E6
N 0.965 ±0.00417
K 5.84E7 ±6.24E6
ΔH -9.176E4 ±689.8
ΔS -272

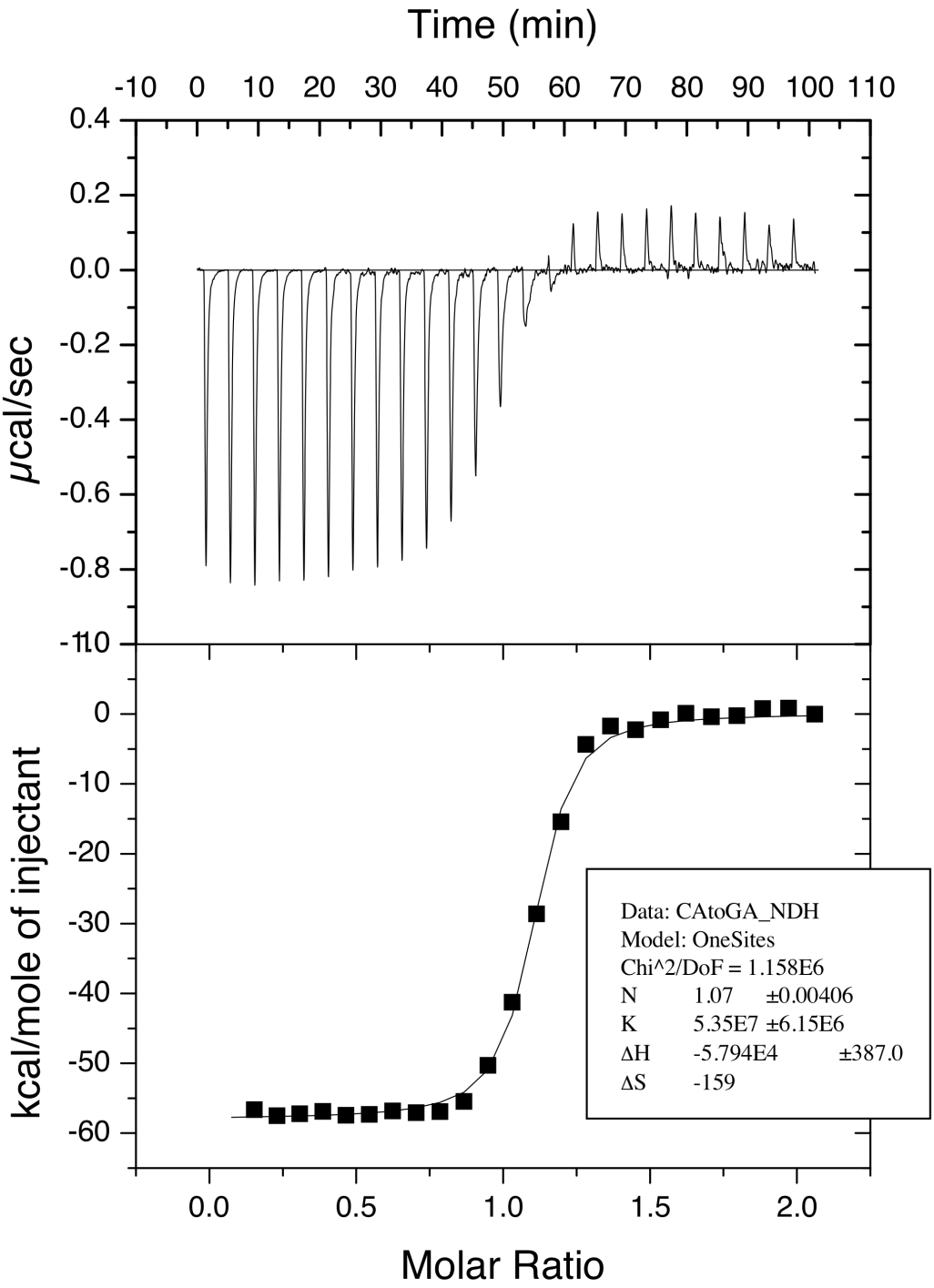
G_C to C_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
µcal/sec
0
-20
-40
-60
-80
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Molar Ratio
Data: GCtoCG_NDH
Model: OneSites
Chi^2/DoF = 8.609E5
N 0.819 ±0.00178
K 4.51E8 ±1.26E8
ΔH -8.156E4 ±429.5
ΔS -234

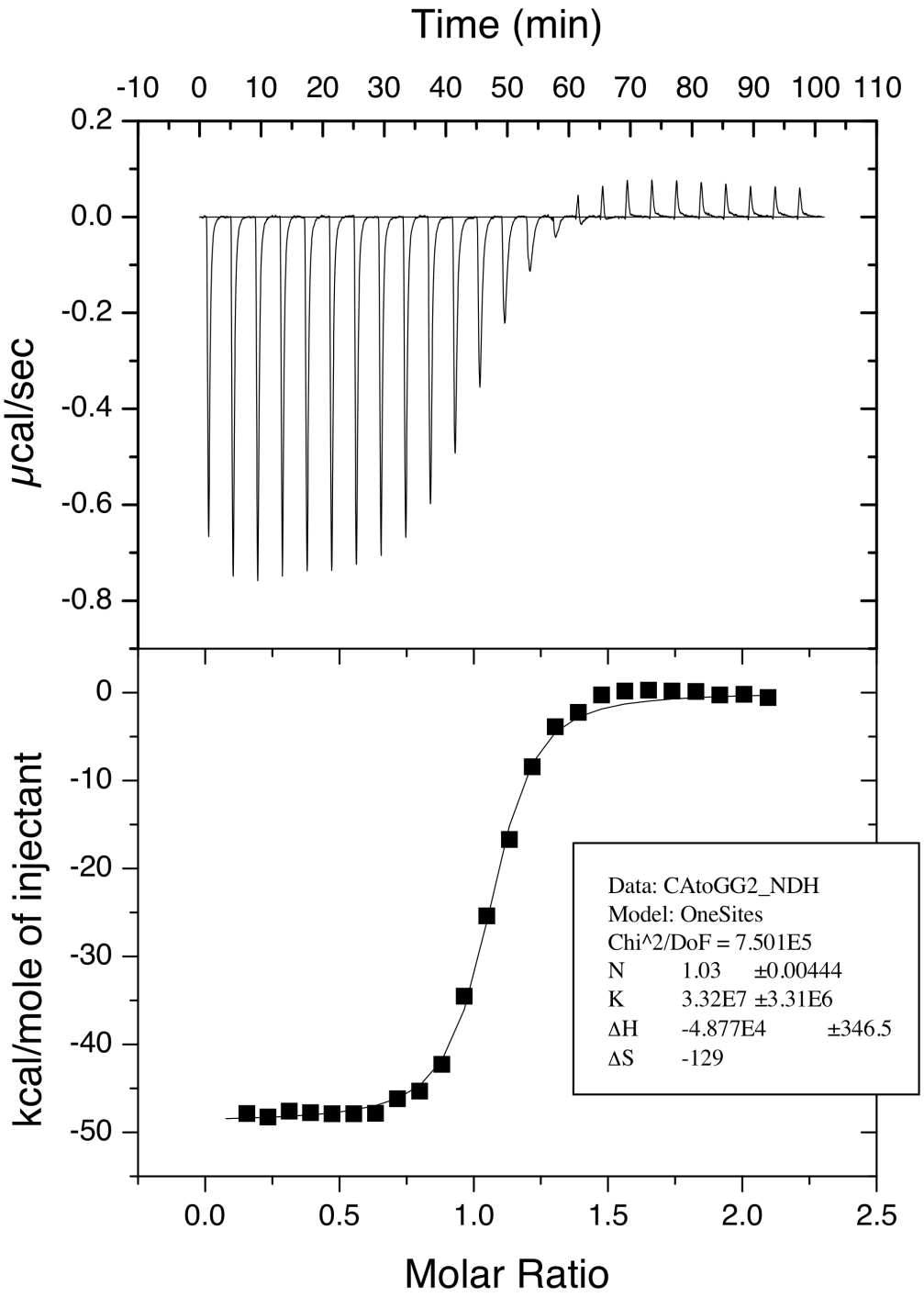
G_C to C_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
µcal/sec
0
-20
-40
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
Molar Ratio
Data: GCtoCAs_NDH
Model: OneSites
Chi^2/DoF = 3.950E5
N 0.880 ±0.00605
K 1.22E7 ±9.17E5
ΔH -6.476E4 ±709.8
ΔS -185

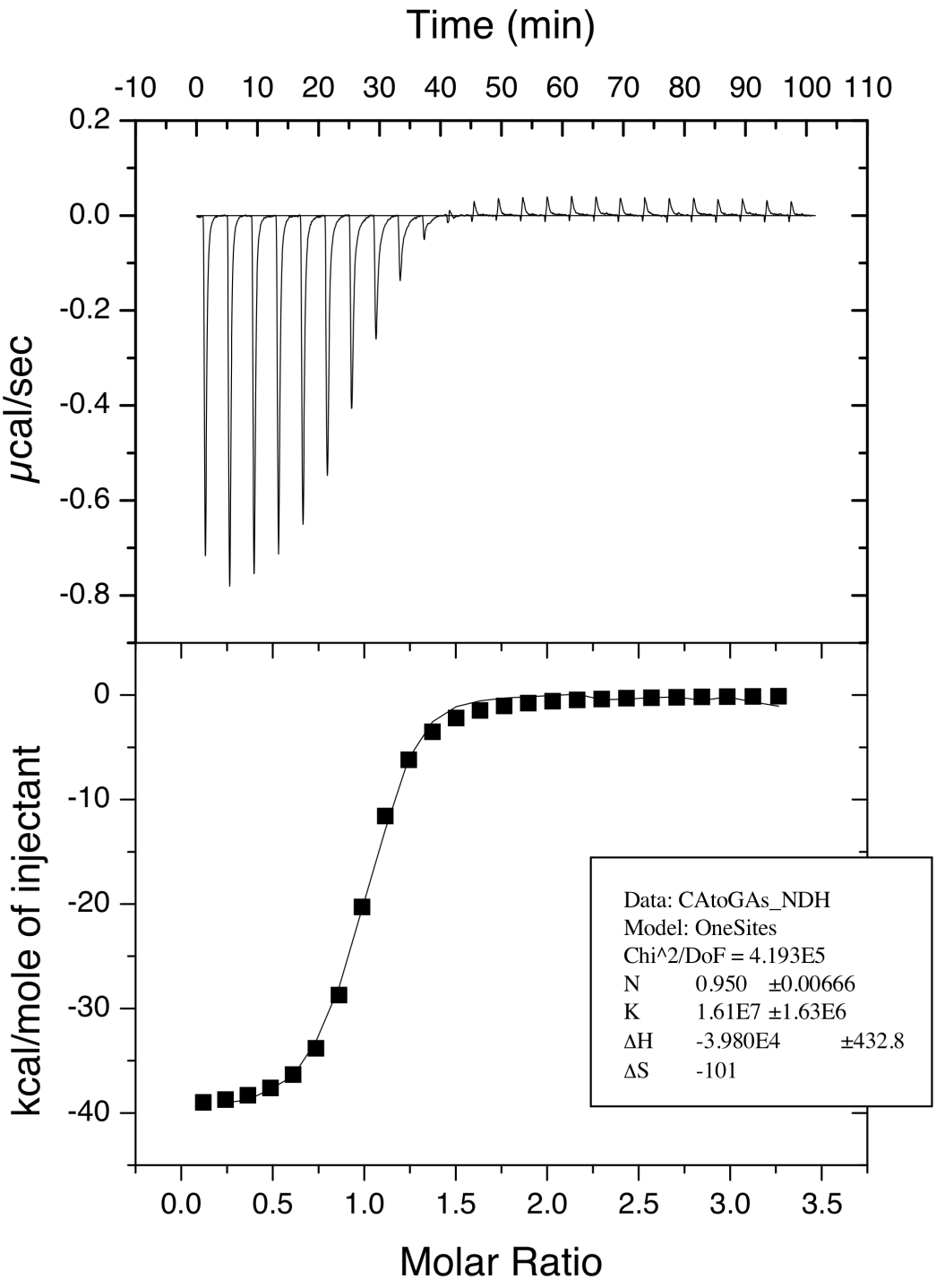
C_A to G_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CAtoGA_NDH
Model: OneSites
Chi^2/DoF = 1.158E6
N 1.07 ±0.00406
K 5.35E7 ±6.15E6
ΔH -5.794E4 ±387.0
ΔS -159

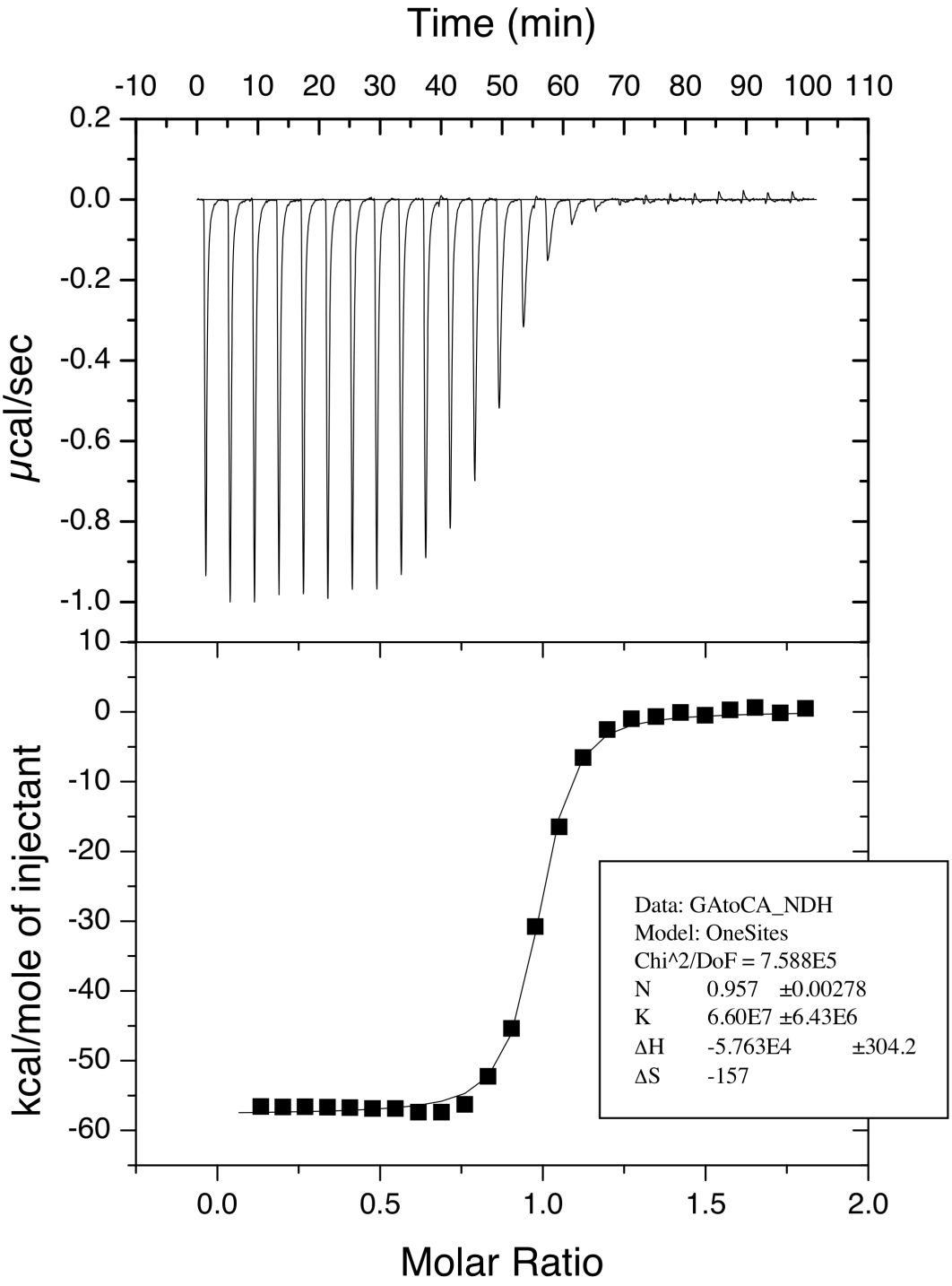
C_A to G_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CAtoGG2_NDH
Model: OneSites
Chi^2/DoF = 7.501E5
N 1.03 ±0.00444
K 3.32E7 ±3.31E6
ΔH -4.877E4 ±346.5
ΔS -129

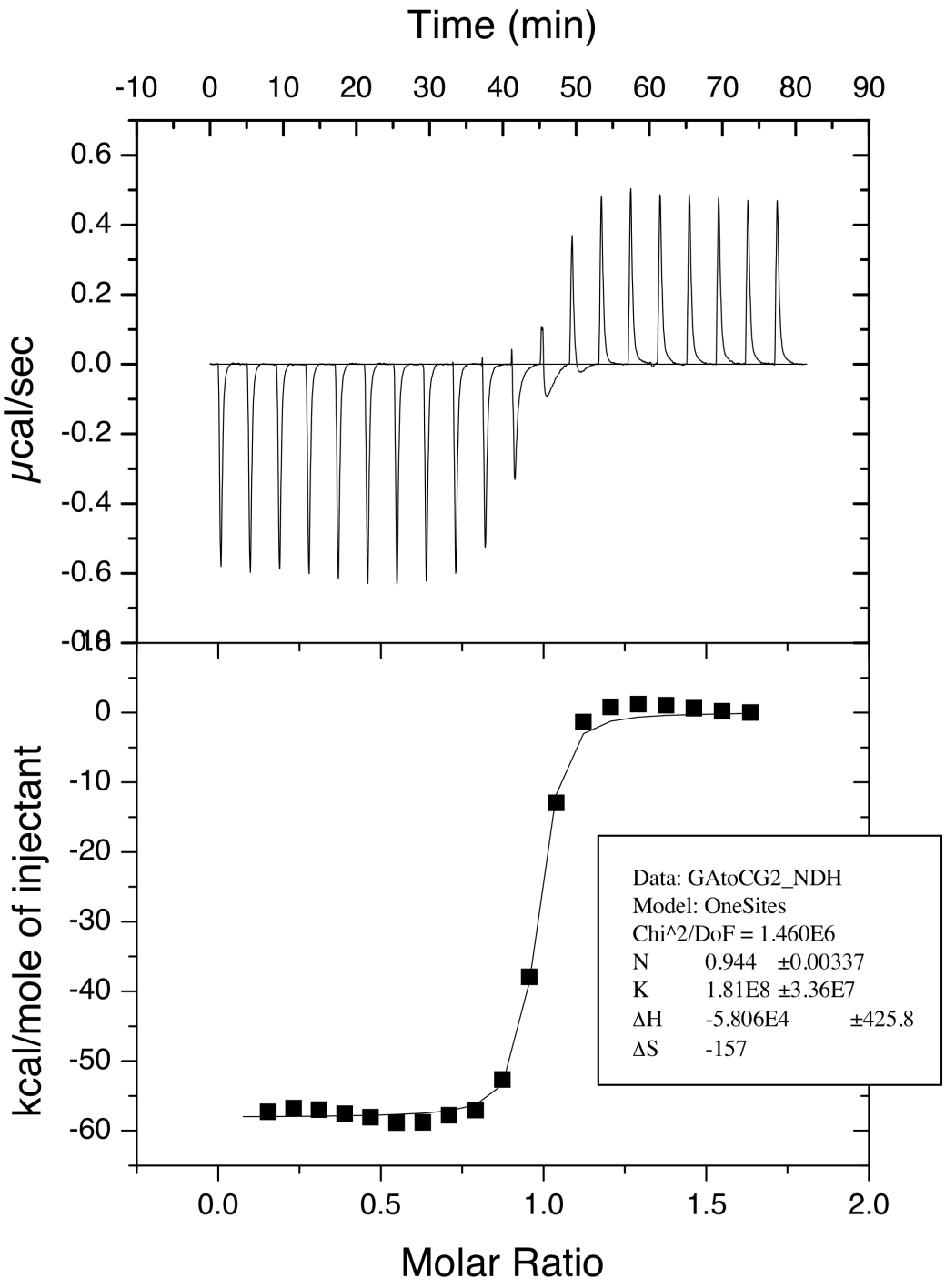
C_A to G_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Molar Ratio
Data: CAtoGAs_NDH
Model: OneSites
Chi^2/DoF = 4.193E5
N 0.950 ±0.00666
K 1.61E7 ±1.63E6
ΔH -3.980E4 ±432.8
ΔS -101

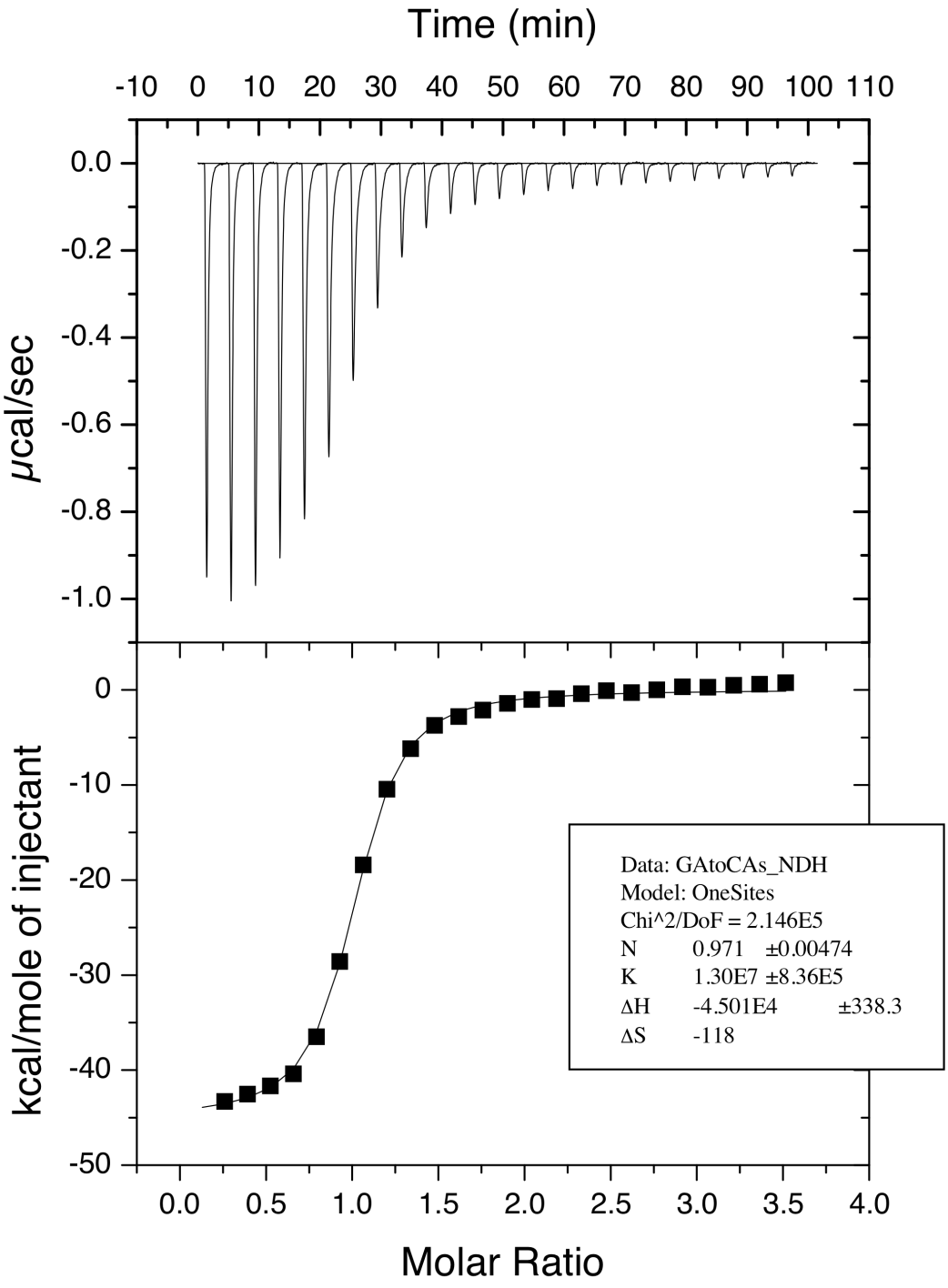
G_A to C_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
10
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GAtoCA_NDH
Model: OneSites
Chi^2/DoF = 7.588E5
N 0.957 ±0.00278
K 6.60E7 ±6.43E6
ΔH -5.763E4 ±304.2
ΔS -157

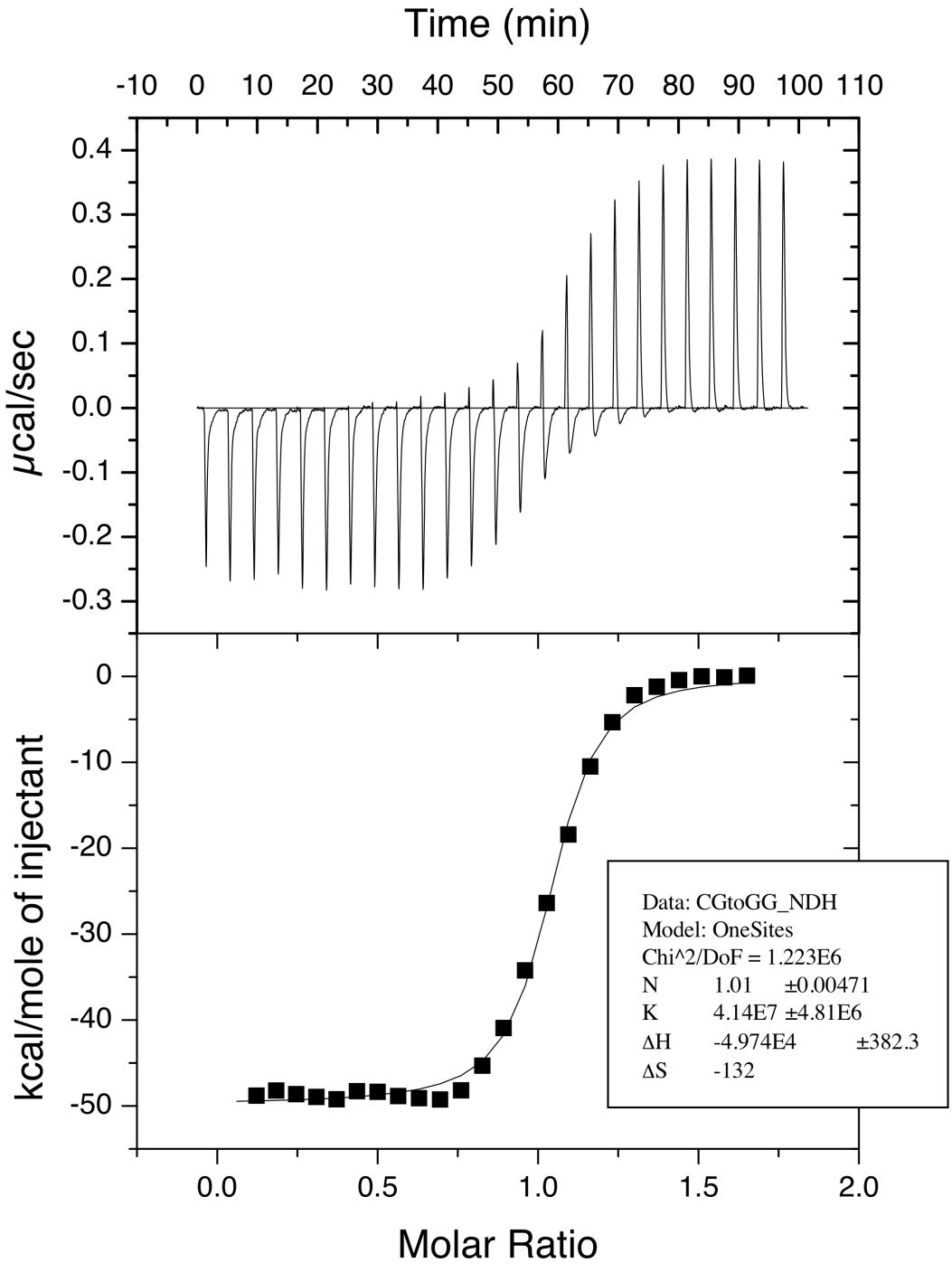
G_A to C_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90
µcal/sec
0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8
kcal/mole of injectant
0 -10 -20 -30 -40 -50 -60
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GAtoCG2_NDH
Model: OneSites
Chi^2/DoF = 1.460E6
N 0.944 ±0.00337
K 1.81E8 ±3.36E7
ΔH -5.806E4 ±425.8
ΔS -157

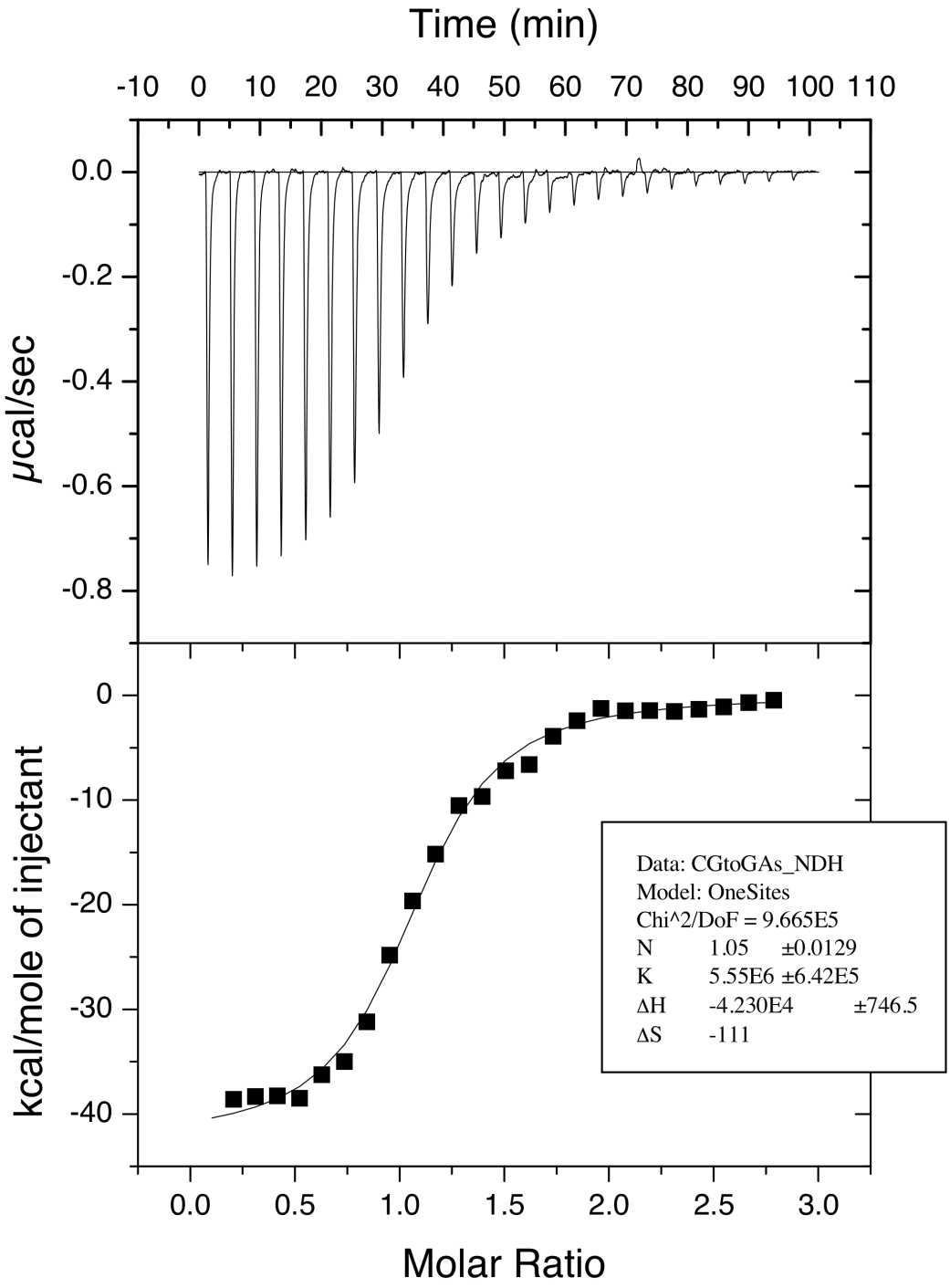
G_A to C_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Molar Ratio
Data: GAtoCAs_NDH
Model: OneSites
Chi^2/DoF = 2.146E5
N 0.971 ±0.00474
K 1.30E7 ±8.36E5
ΔH -4.501E4 ±338.3
ΔS -118

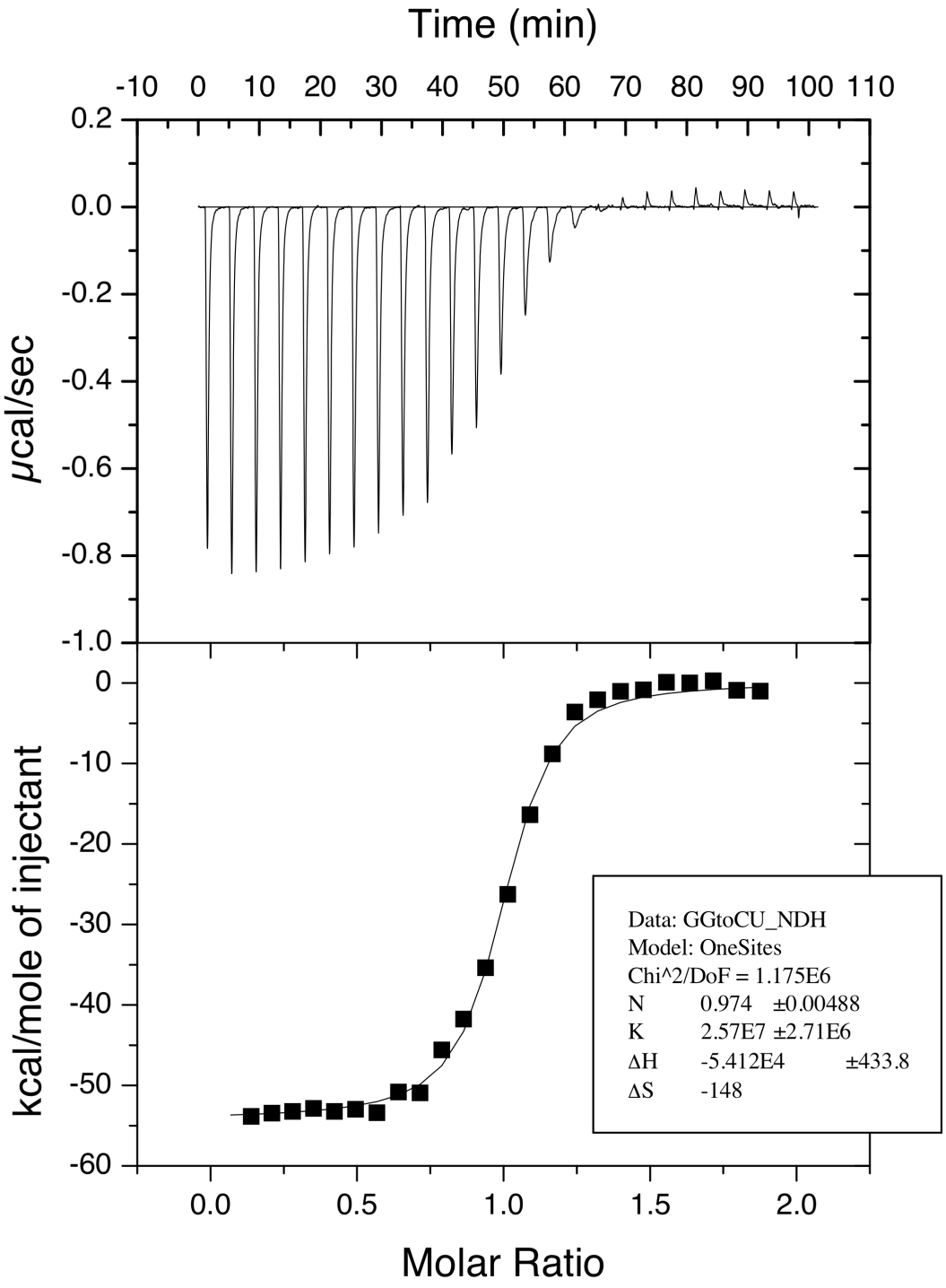
C_G to G_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
µcal/sec
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
kcal/mole of injectant
0
-10
-20
-30
-40
-50
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CGtoGG_NDH
Model: OneSites
Chi^2/DoF = 1.223E6
N 1.01 ±0.00471
K 4.14E7 ±4.81E6
ΔH -4.974E4 ±382.3
ΔS -132

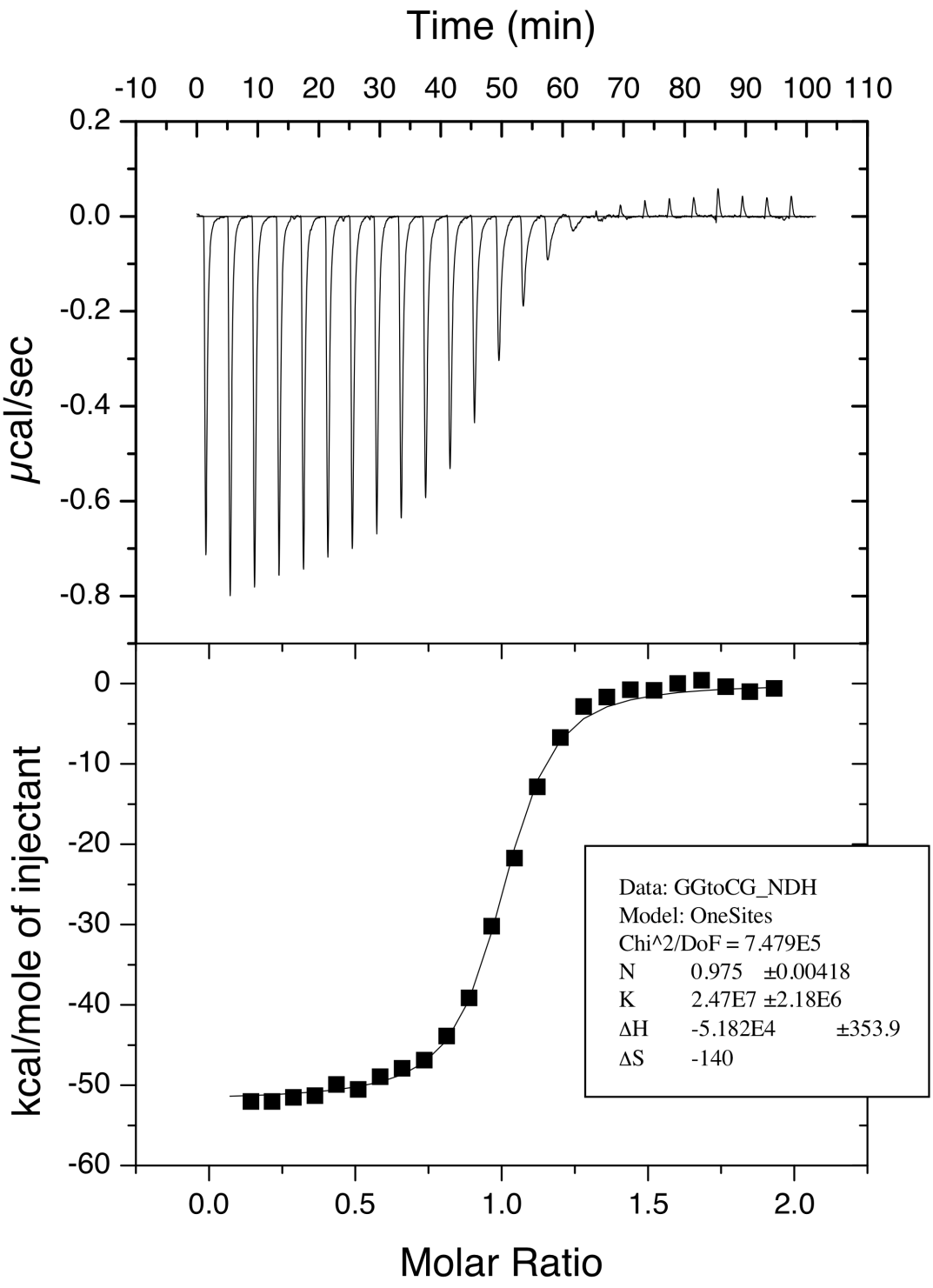
C_G to G_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Molar Ratio
Data: CGtoGAs_NDH
Model: OneSites
Chi^2/DoF = 9.665E5
N 1.05 ±0.0129
K 5.55E6 ±6.42E5
ΔH -4.230E4 ±746.5
ΔS -111

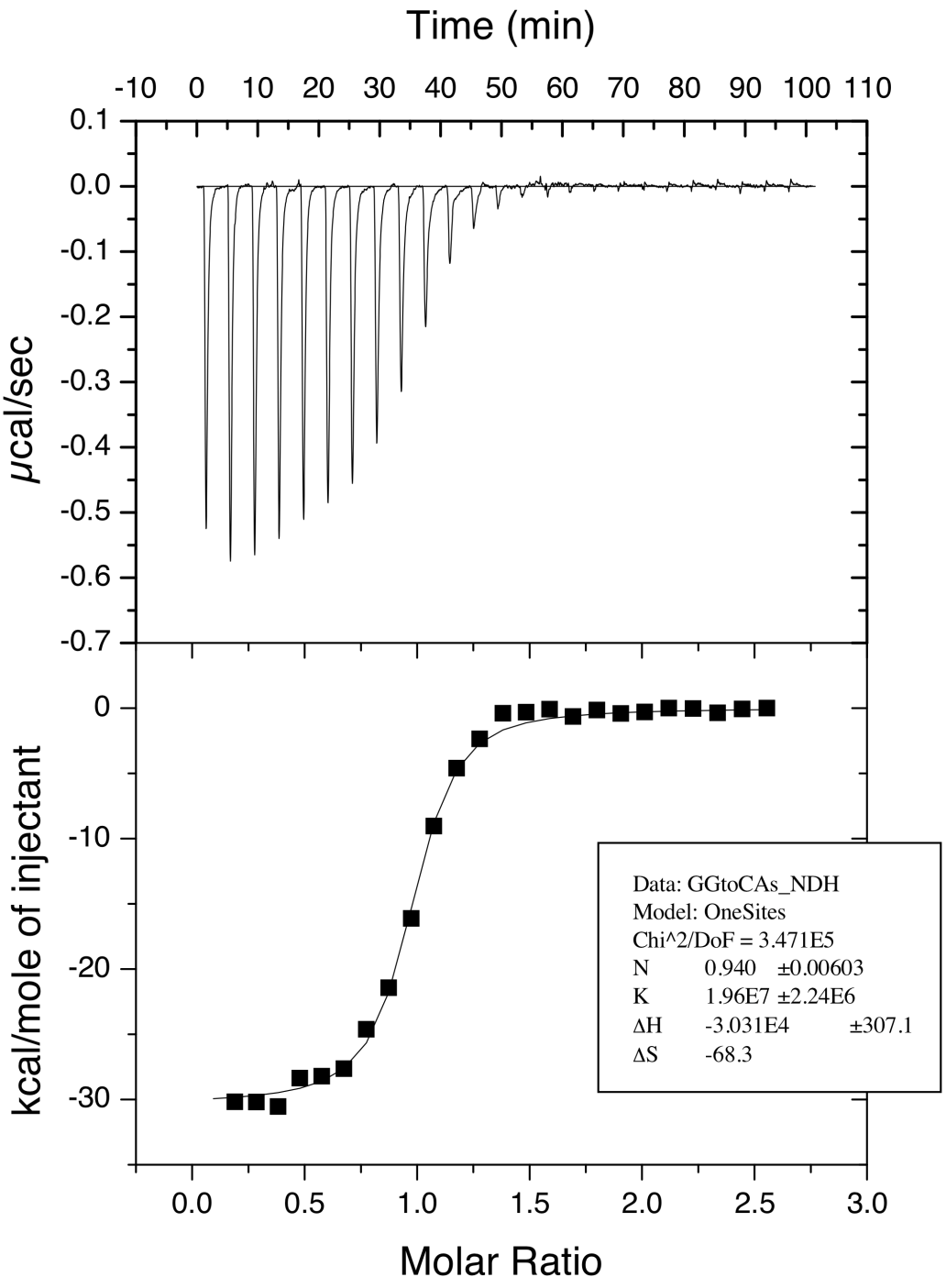
G_G to C_U


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
µcal/sec
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GGtoCU_NDH
Model: OneSites
Chi^2/DoF = 1.175E6
N 0.974 ±0.00488
K 2.57E7 ±2.71E6
ΔH -5.412E4 ±433.8
ΔS -148

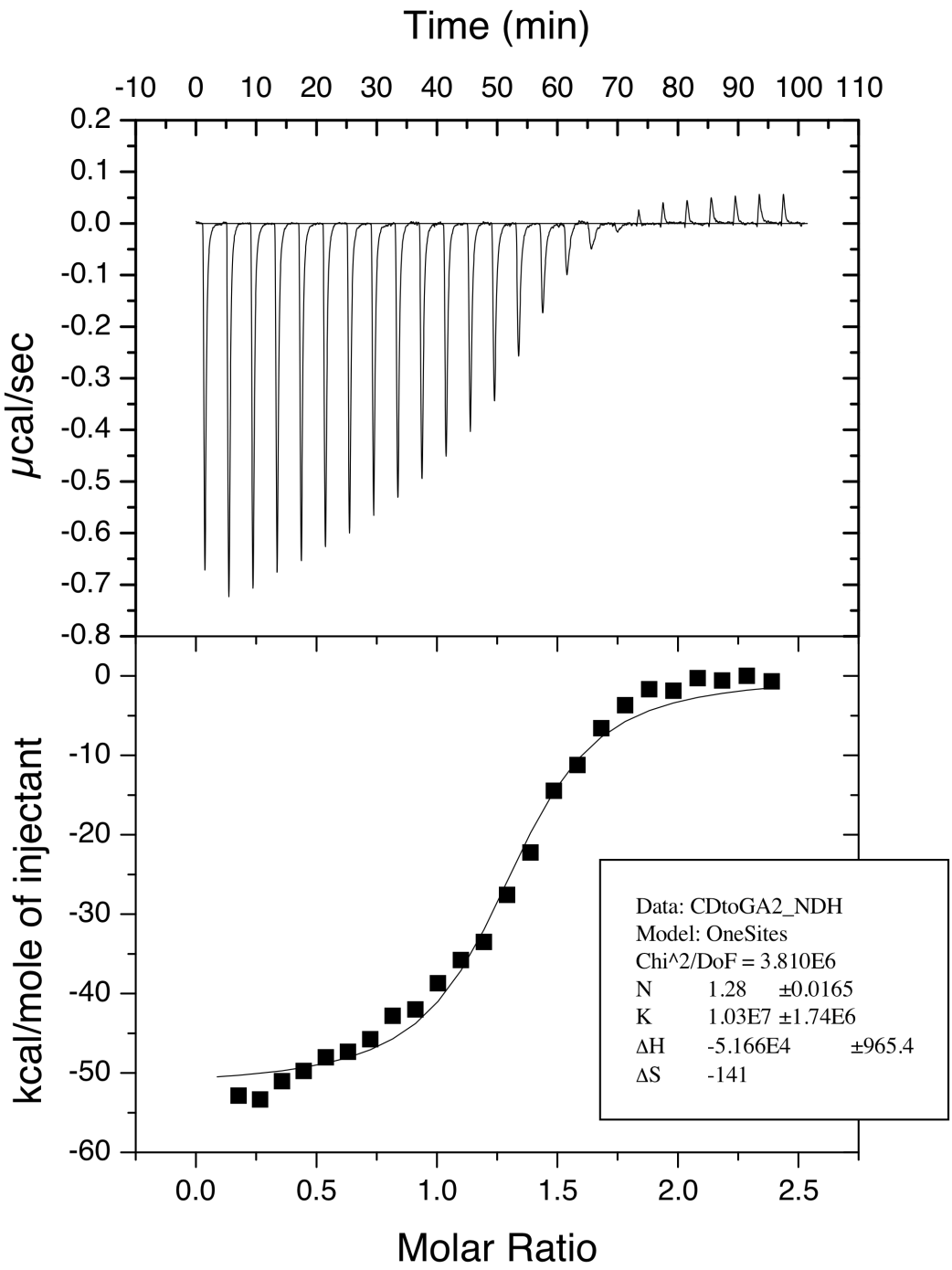
G_G to C_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GGtoCG_NDH
Model: OneSites
Chi^2/DoF = 7.479E5
N 0.975 ±0.00418
K 2.47E7 ±2.18E6
ΔH -5.182E4 ±353.9
ΔS -140

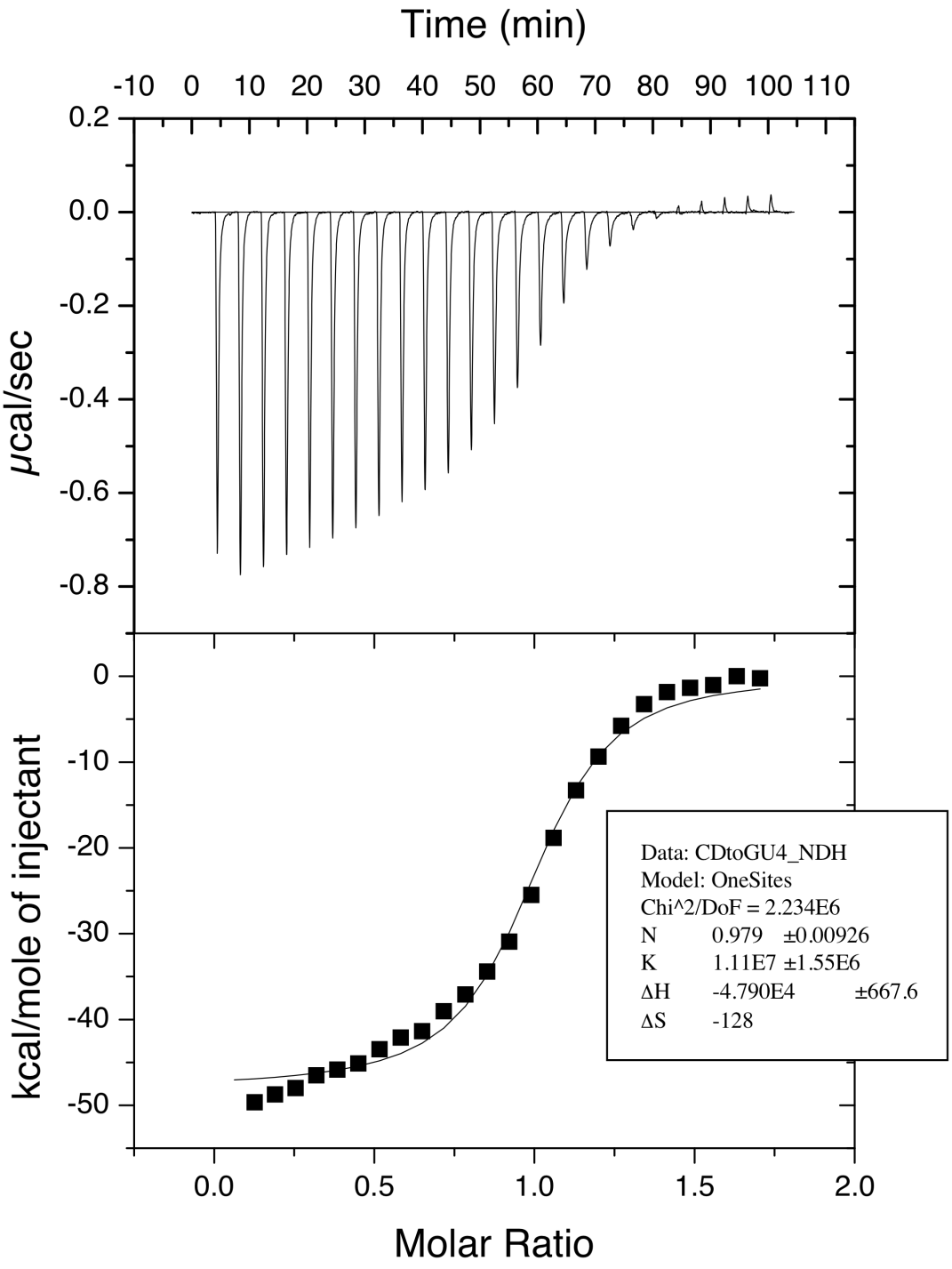
G_G to C_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
µcal/sec
0
-10
-20
-30
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Molar Ratio
Data: GGtoCAs_NDH
Model: OneSites
Chi^2/DoF = 3.471E5
N 0.940 ±0.00603
K 1.96E7 ±2.24E6
ΔH -3.031E4 ±307.1
ΔS -68.3

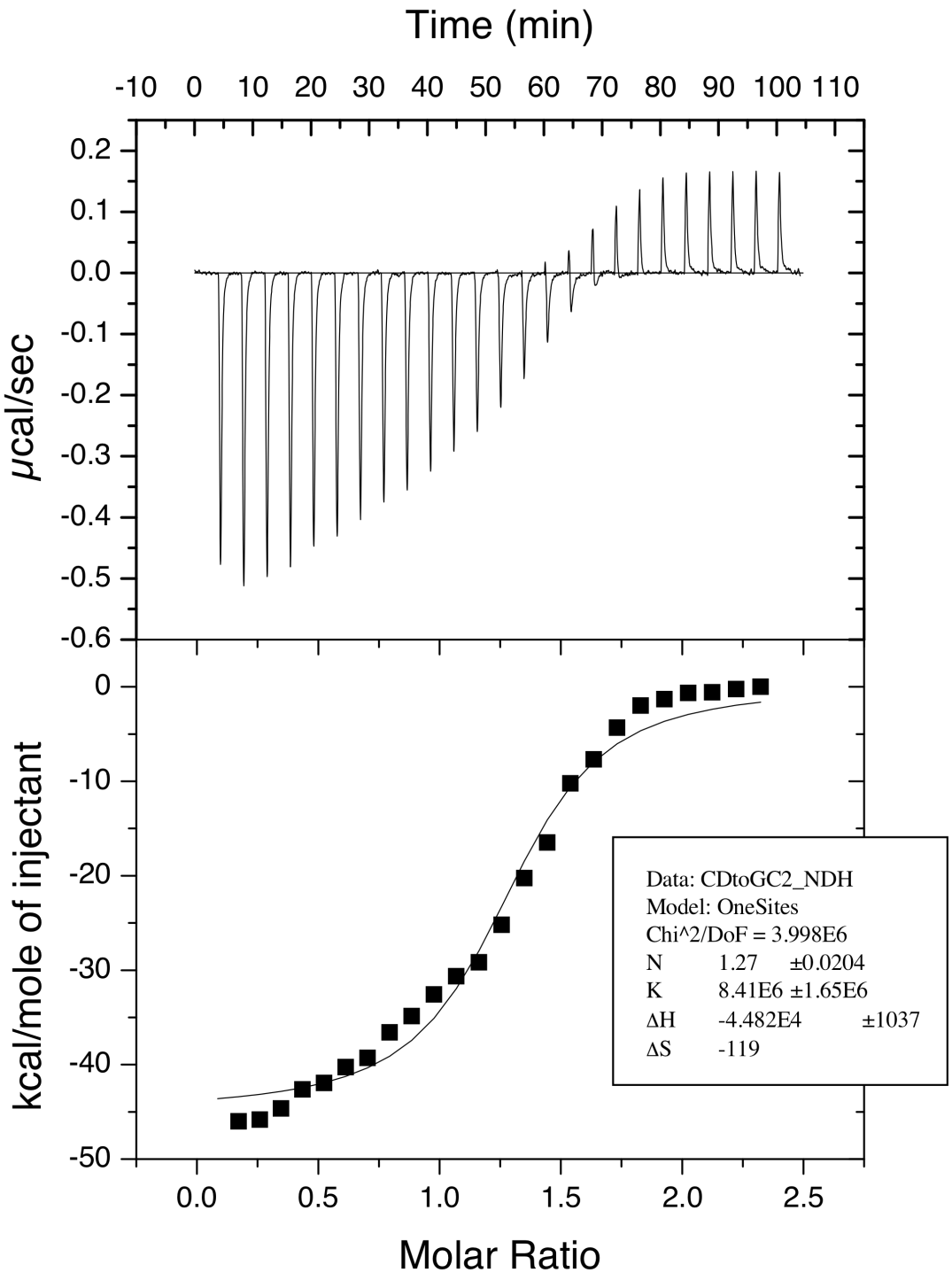
C_D to G_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CDtoGA2_NDH
Model: OneSites
Chi^2/DoF = 3.810E6
N 1.28 ±0.0165
K 1.03E7 ±1.74E6
ΔH -5.166E4 ±965.4
ΔS -141

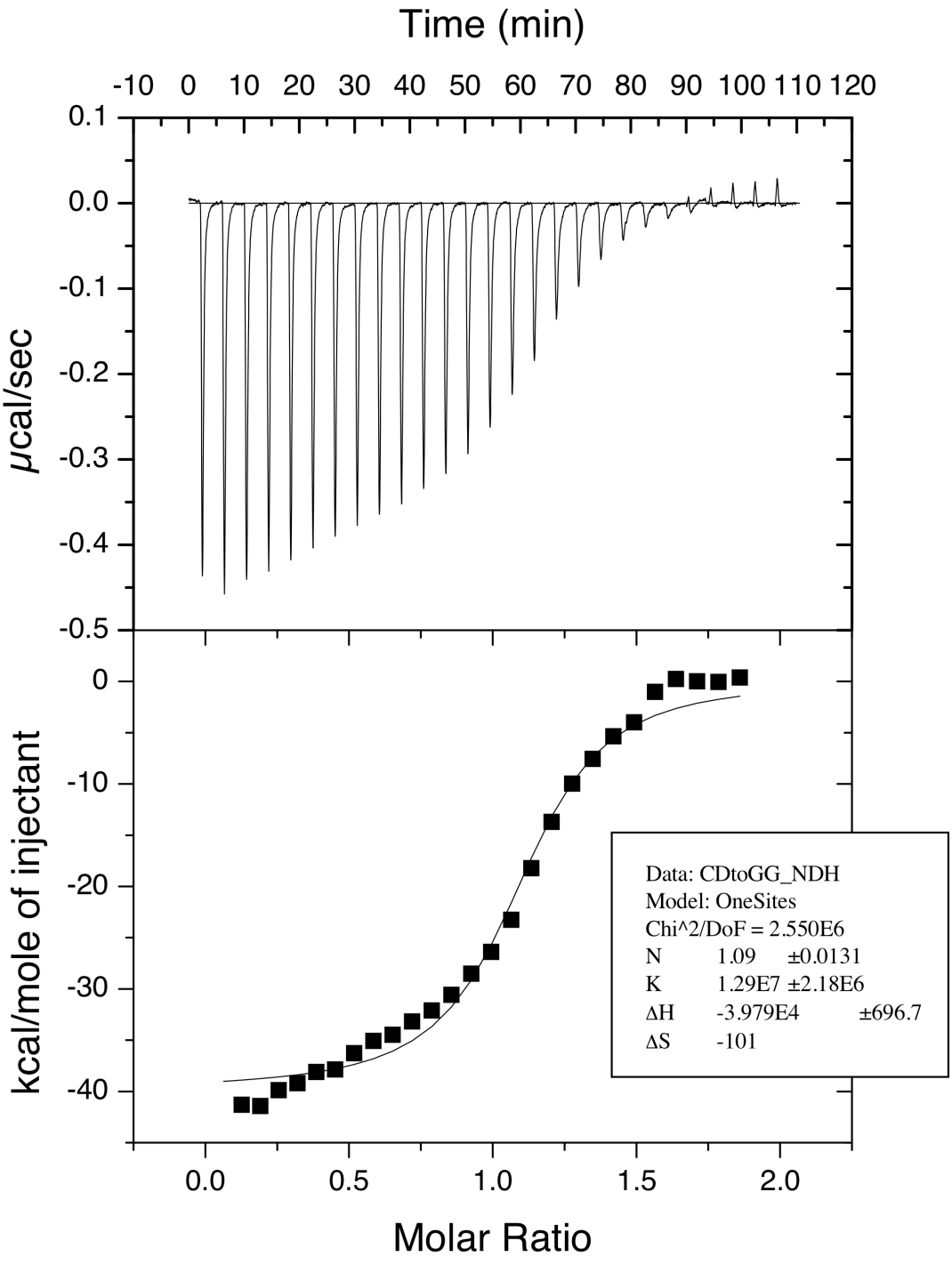
C_D to G_U


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CDtoGU4_NDH
Model: OneSites
Chi^2/DoF = 2.234E6
N 0.979 ±0.00926
K 1.11E7 ±1.55E6
ΔH -4.790E4 ±667.6
ΔS -128

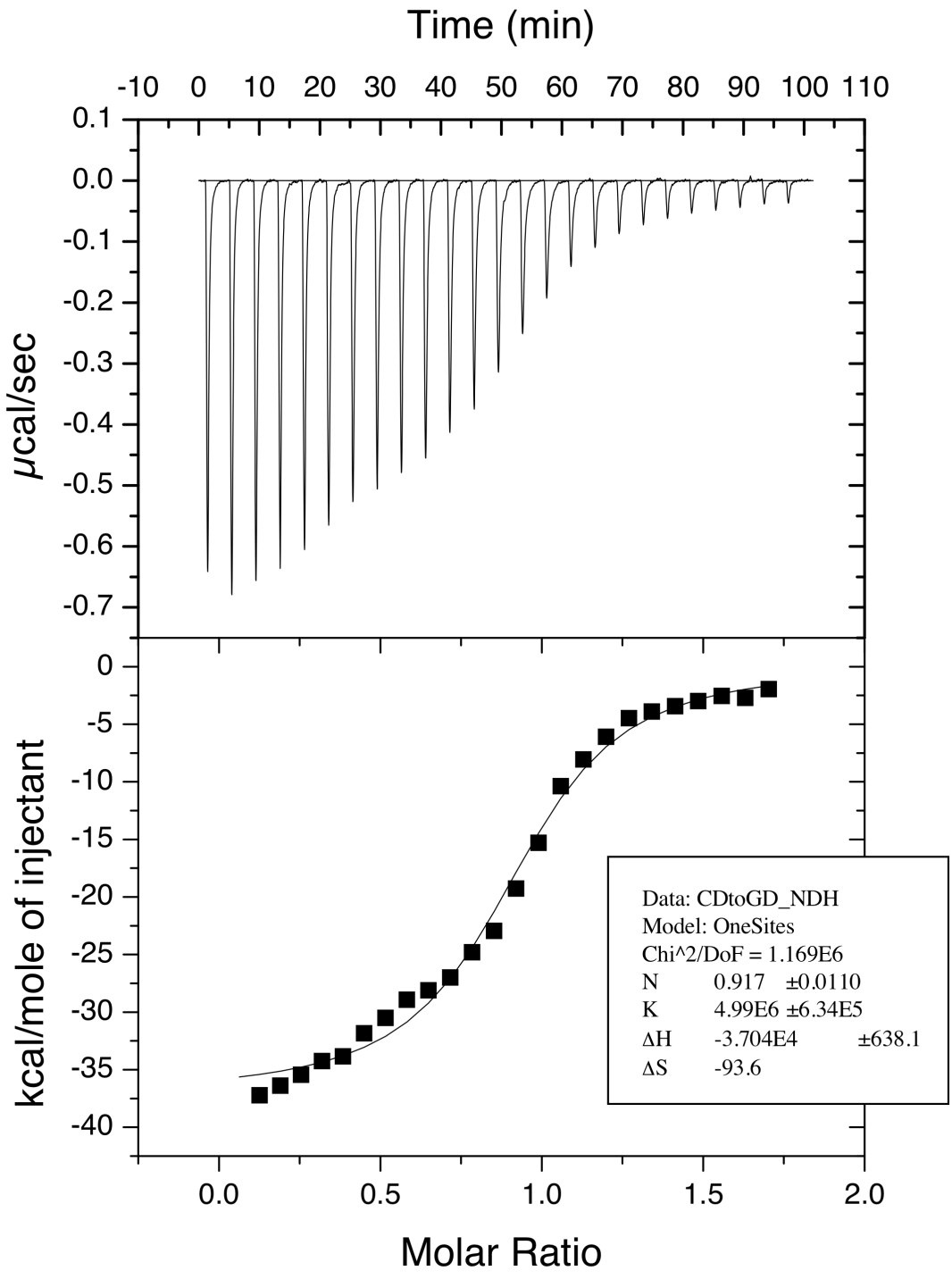
C_D to G_C


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: CDtoGC2_NDH
Model: OneSites
Chi^2/DoF = 3.998E6
N 1.27 ±0.0204
K 8.41E6 ±1.65E6
ΔH -4.482E4 ±1037
ΔS -119

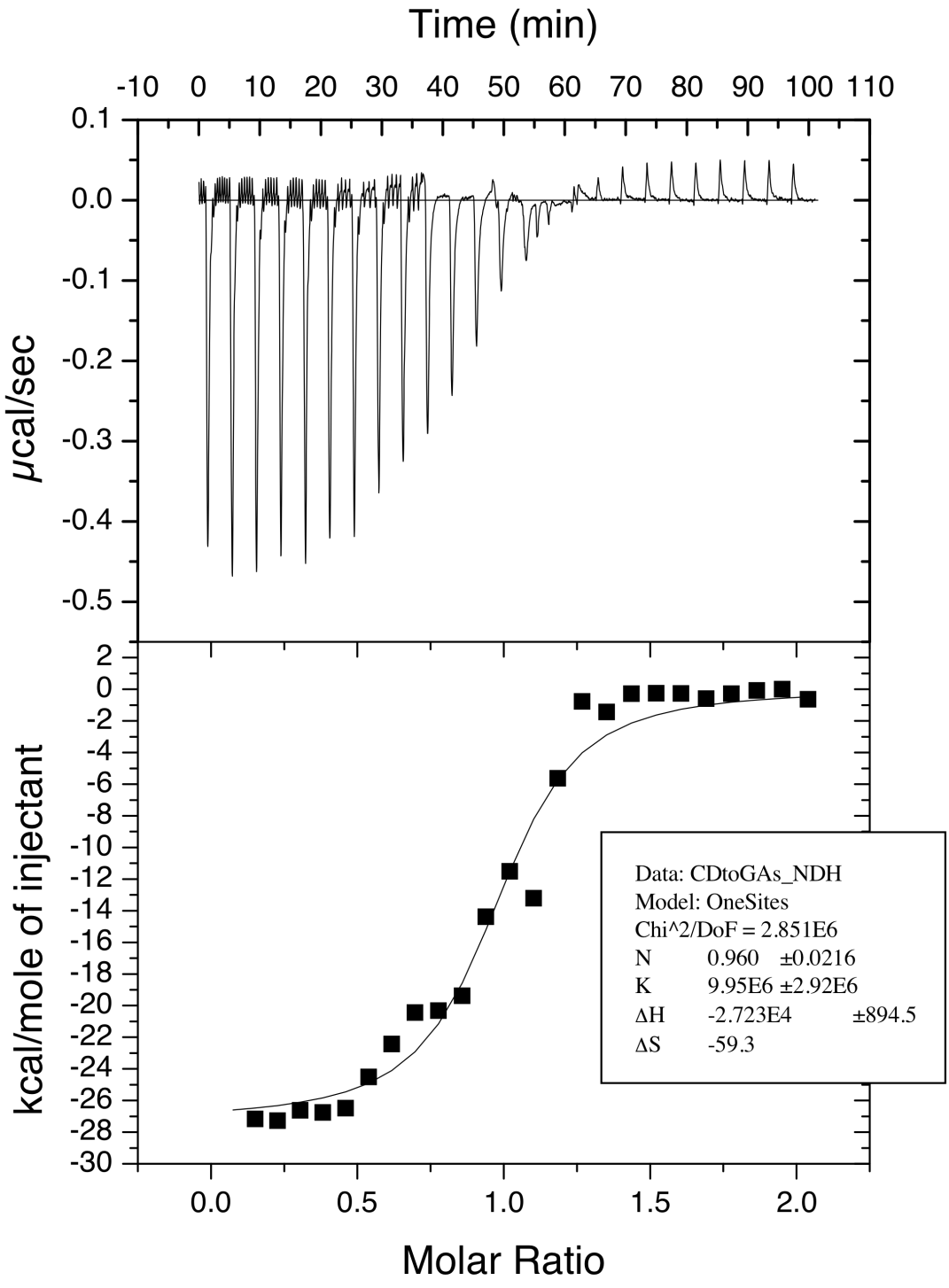
C_D to G_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110 120
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
µcal/sec
0
-10
-20
-30
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CDtoGG_NDH
Model: OneSites
Chi^2/DoF = 2.550E6
N 1.09 ±0.0131
K 1.29E7 ±2.18E6
ΔH -3.979E4 ±696.7
ΔS -101

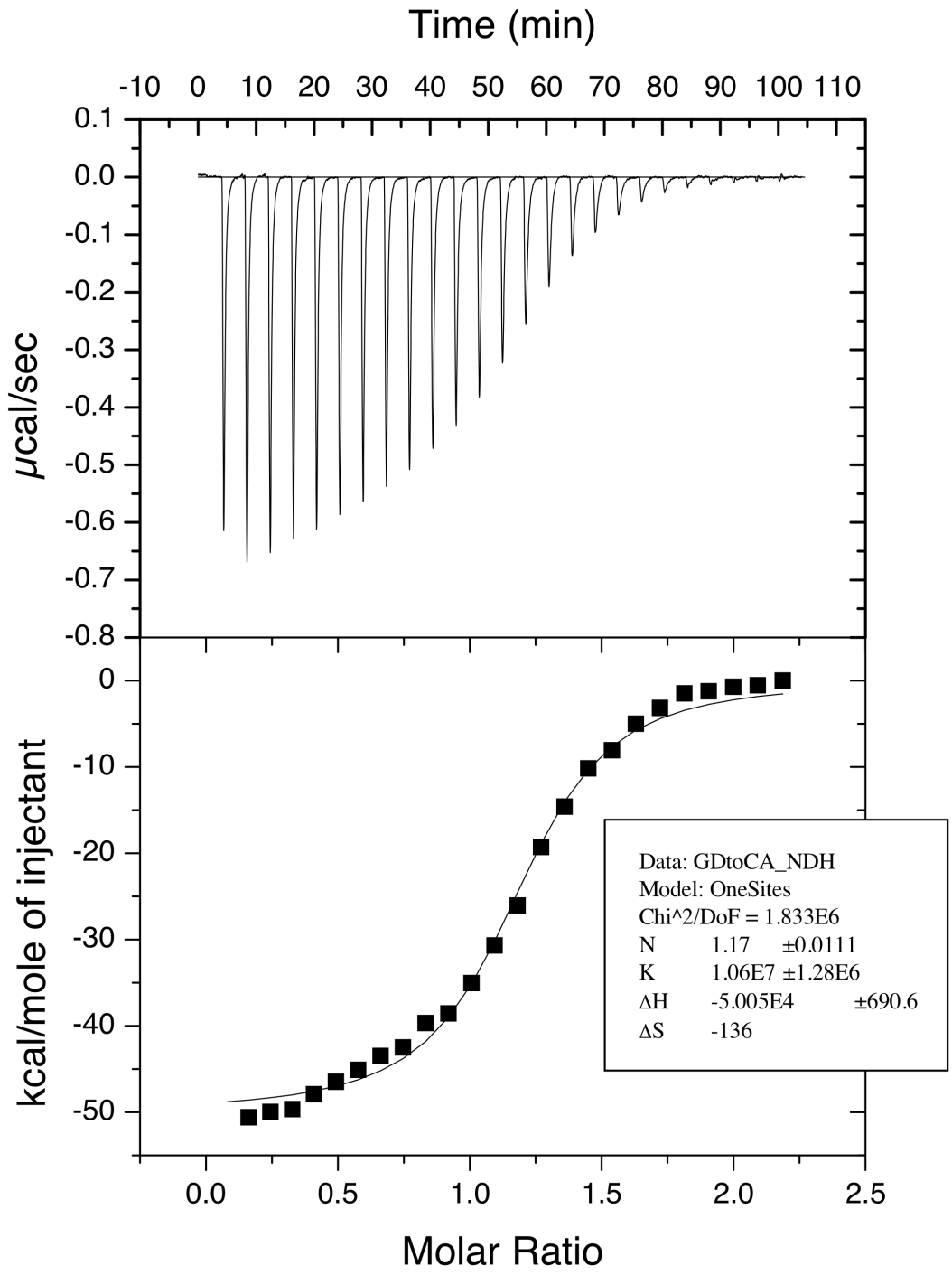
C_D to G_D


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
µcal/sec
0
-5
-10
-15
-20
-25
-30
-35
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CDtoGD_NDH
Model: OneSites
Chi^2/DoF = 1.169E6
N 0.917 ±0.0110
K 4.99E6 ±6.34E5
ΔH -3.704E4 ±638.1
ΔS -93.6

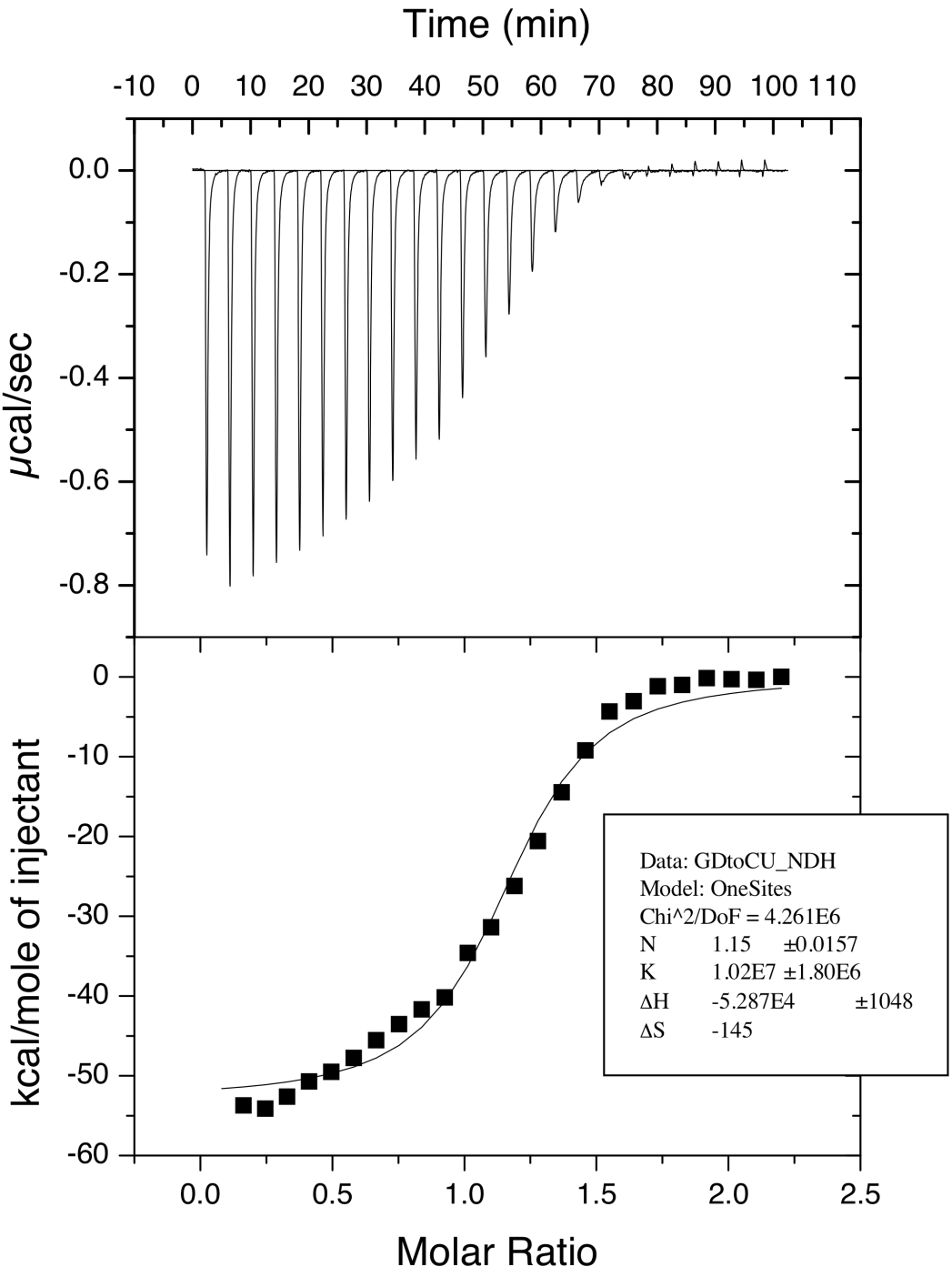
C_D to G_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1 0.0 -0.1 -0.2 -0.3 -0.4 -0.5
µcal/sec
2 0 -2 -4 -6 -8 -10 -12 -14 -16 -18 -20 -22 -24 -26 -28 -30
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: CDtoGAs_NDH
Model: OneSites
Chi^2/DoF = 2.851E6
N 0.960 ±0.0216
K 9.95E6 ±2.92E6
ΔH -2.723E4 ±894.5
ΔS -59.3

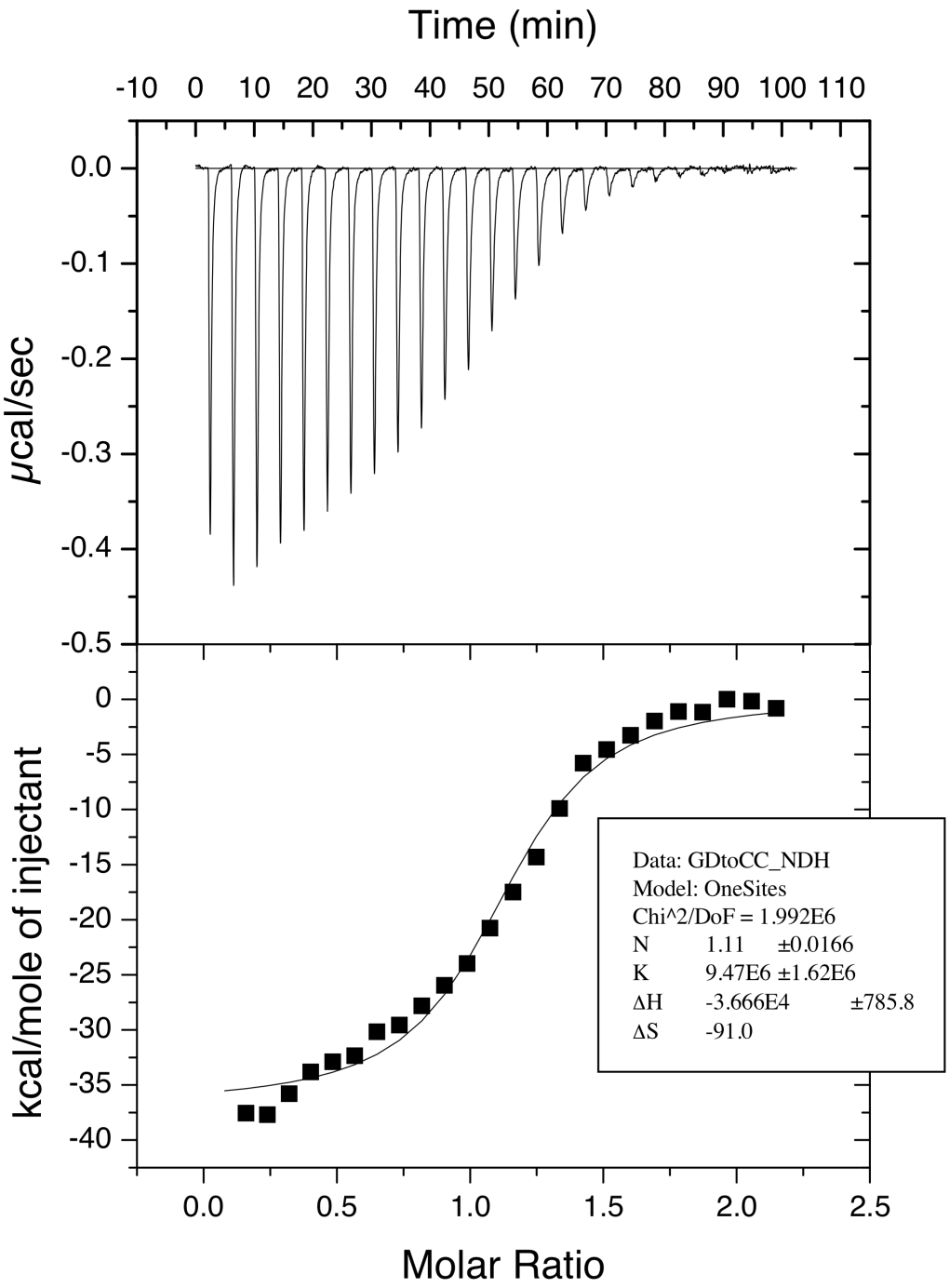
G_D to C_A


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
µcal/sec
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
kcal/mole of injectant
0
-10
-20
-30
-40
-50
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: GDtoCA_NDH
Model: OneSites
Chi^2/DoF = 1.833E6
N 1.17 ±0.0111
K 1.06E7 ±1.28E6
ΔH -5.005E4 ±690.6
ΔS -136

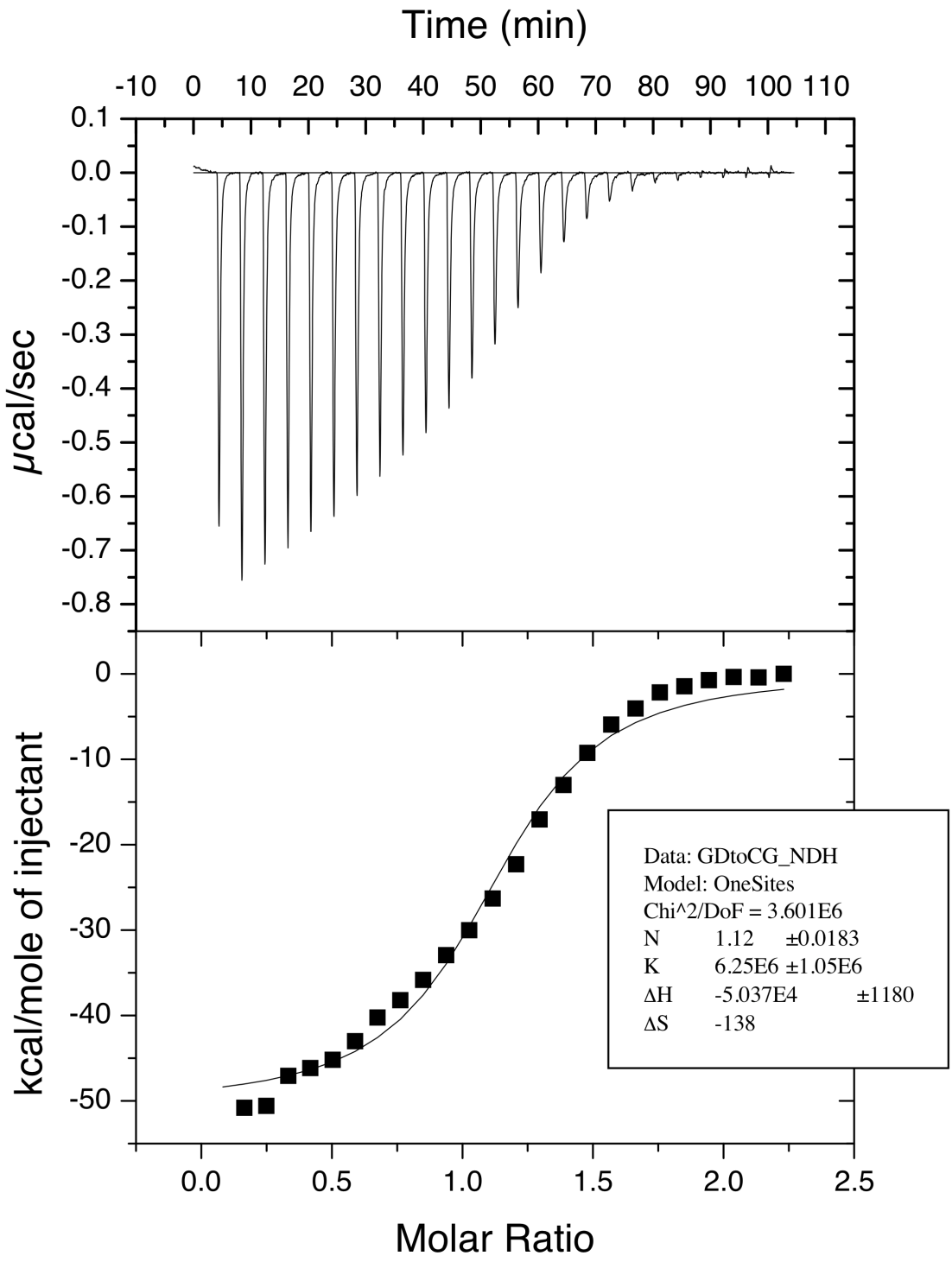
G_D to C_U


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.2
-0.4
-0.6
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
-60
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: GDtoCU_NDH
Model: OneSites
Chi^2/DoF = 4.261E6
N 1.15 ±0.0157
K 1.02E7 ±1.80E6
ΔH -5.287E4 ±1048
ΔS -145

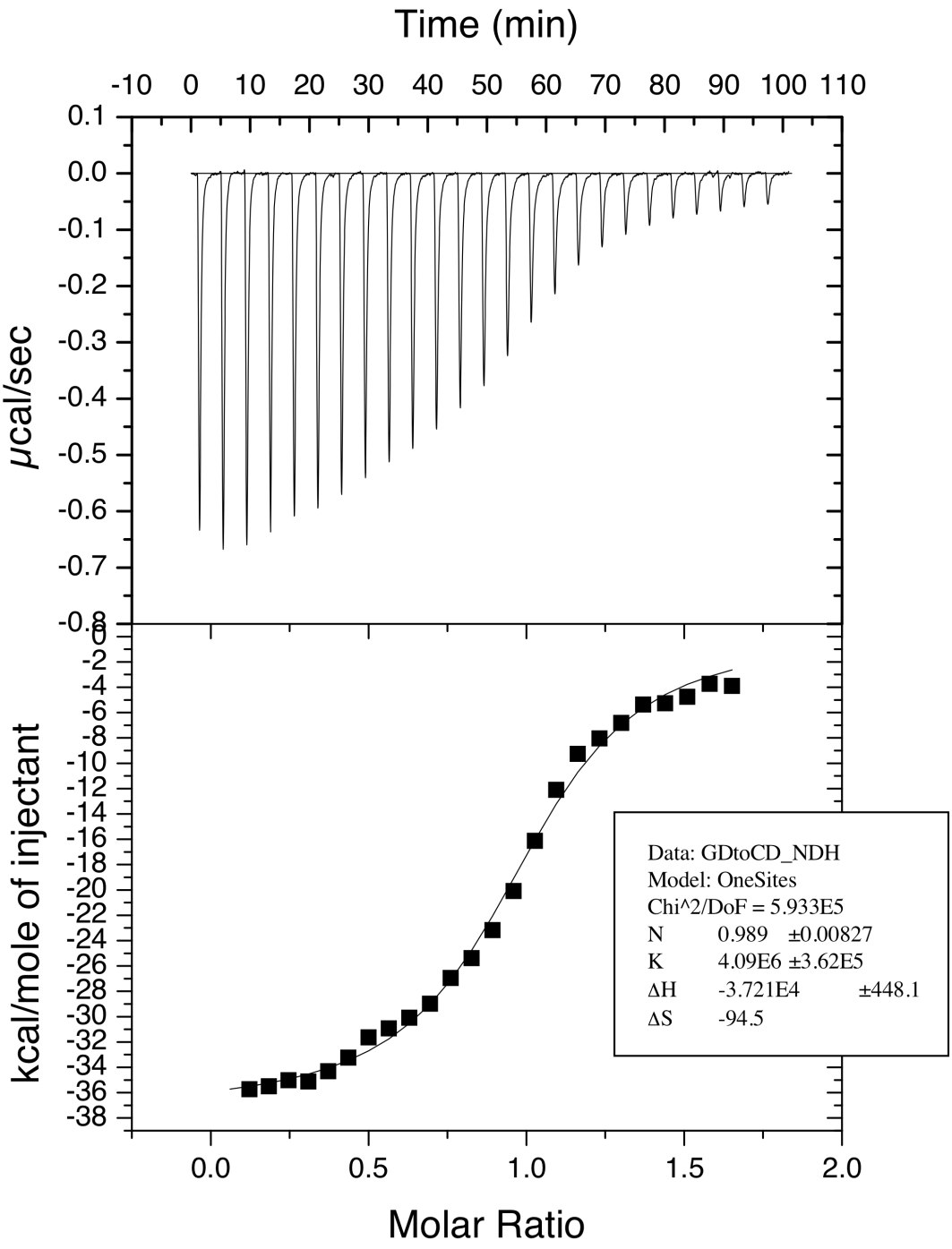
G_D to C_C


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
µcal/sec
0
-5
-10
-15
-20
-25
-30
-35
-40
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: GDtoCC_NDH
Model: OneSites
Chi^2/DoF = 1.992E6
N 1.11 ±0.0166
K 9.47E6 ±1.62E6
ΔH -3.666E4 ±785.8
ΔS -91.0

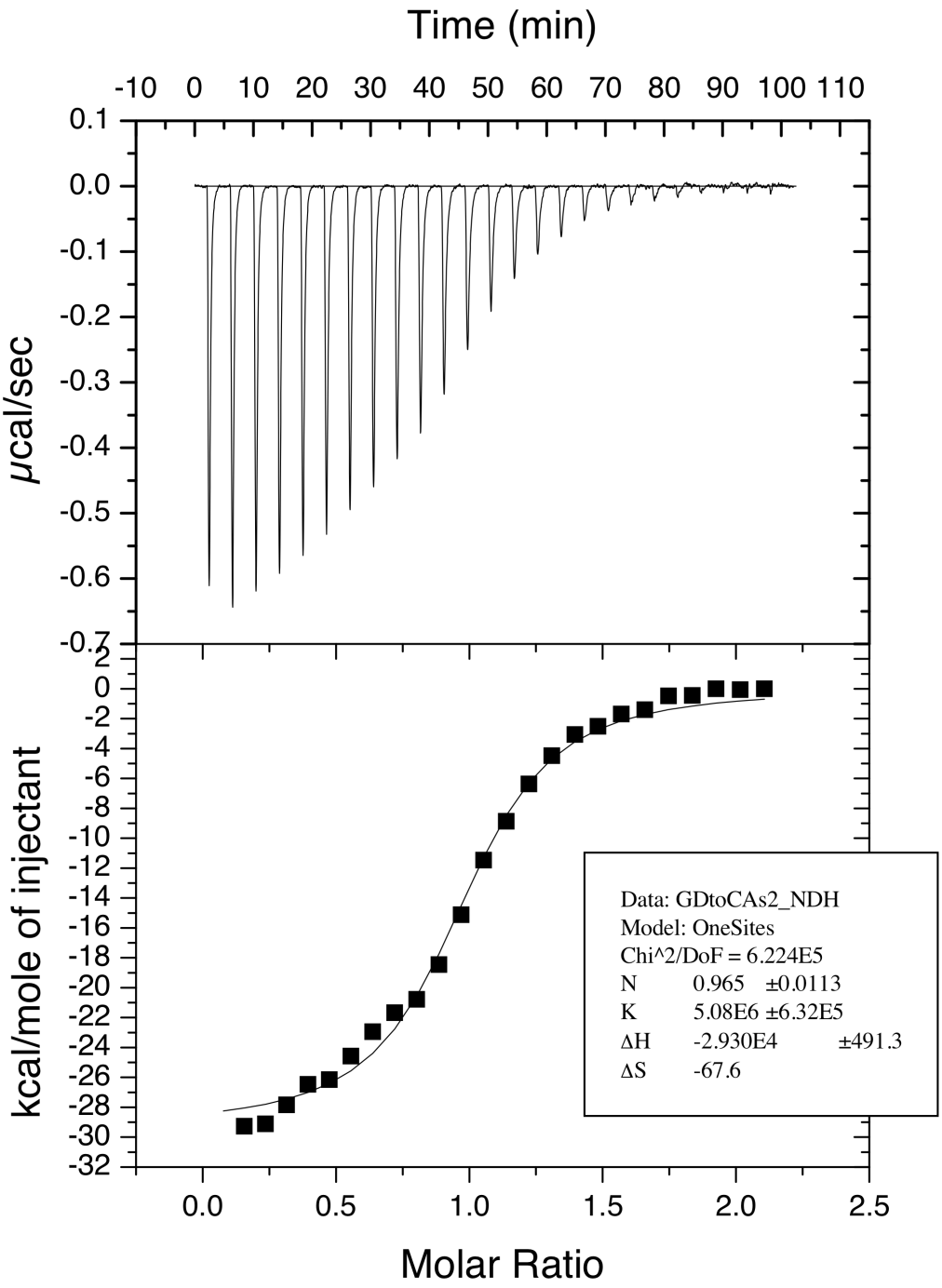
G_D to C_G


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
µcal/sec
0
-10
-20
-30
-40
-50
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: GDtoCG_NDH
Model: OneSites
Chi^2/DoF = 3.601E6
N 1.12 ±0.0183
K 6.25E6 ±1.05E6
ΔH -5.037E4 ±1180
ΔS -138

G_D to C_D


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
µcal/sec
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
-34
-36
-38
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0
Molar Ratio
Data: GDtoCD_NDH
Model: OneSites
Chi^2/DoF = 5.933E5
N 0.989 ±0.00827
K 4.09E6 ±3.62E5
ΔH -3.721E4 ±448.1
ΔS -94.5

G_D to C_As


Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100 110
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
µcal/sec
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
kcal/mole of injectant
0.0 0.5 1.0 1.5 2.0 2.5
Molar Ratio
Data: GDtoCAs2_NDH
Model: OneSites
Chi^2/DoF = 6.224E5
N 0.965 ±0.0113
K 5.08E6 ±6.32E5
ΔH -2.930E4 ±491.3
ΔS -67.6

Acknowledgements

PROF. JAY S. SIEGEL

for giving me the opportunity to work in his group, and for his excellent guidance and support.

PROF. KIM K. BALDRIDGE

for being in my committee as well as for helpful discussions.

PROF. ROLAND K. O. SIGEL

for being in my committee and for help with the pK_a measurements.

DR. GILLES BRUYLANTS AND PROF. KRISTIN BARTIK

for enormous help with the ITC measurements and for their hospitality.

DR. BERNHARD SPINGLER

for help with the ITC measurements.

PROF. NATHAN LUEDTKE

for help with everything related to DNA chemistry.

NADJA BROSS AND SIMON JURT

for friendly help with my NMR questions.

DR. LAURENT BIGLER AND THE WHOLE MS TEAM

for enormous help with all my MS needs.

DR. BENNO BISCHOF

for all his help in the beginning of my PhD, for a great teamwork and friendship.

DR. YAZMIN M. OSORNIO AND DR. MAREK KOMISARSKI

for synthetic ideas.

PETER ÜBELHART

for help with the HPLC and general synthetic advices.

ANGELA STEINAUER

for fantastic team work and a great friendship.

FLORIAN WEISSHAR

for great synthetic contributions towards crystallization of our compounds.

ANNA M. BUTTERFIELD AND FABIENNE FURRER

for the good writing atmosphere, the friendship and essential proof-reading.

ALL THE SIEGEL AND FINNEY GROUP MEMBERS

for the good lab atmosphere and all the chemical and non-chemical discussions.

THE UNIVERSITY OF ZURICH AND SNF

for funding.

CHRISTA AND WERNER ROOST, SANDRA, PATRICK, MICHAEL SIGL AND MIRJAM SPREITER

for all their support, encouragement and love.

Caroline Roost

Schwamendingenstrasse 16, 8050 Zürich

Phone +41 (0) 44 311 68 79 E-mail roost@oci.uzh.ch

Education

Apr 2008 - present	PhD with Prof. Dr. J. S. Siegel, University of Zurich *"Synthesis of a C-Deoxyribonucleoside Modelled on Diaminopyrimidine: Evaluation of its Compatibility with the Modern DNA Alphabet"*
Sep 2006 – Dec 2007	Master Studies in Chemistry and Molecular Science with main focus on Chemical Biology and Organic Synthesis, University of Berne Master Thesis with Prof. Dr. R. Häner (Grade: 5.5): *"Pyridinium Pyrenes: Synthesis, Spectroscopic Properties and DNA Quadruplex Binding"*
Sep 2003 – Jun 2006	Bachelor Studies in Chemistry and Molecular Science, University of Berne Bachelor Thesis with Dr. S. Schürch (Grade: 5.5): *"Electrospray Ionisation Mass Spectrometric Study of Magnesium-Oligonucleotides and Cisplatin-Oligonucleotides Complexes*"
Sep 2001 – Jun 2003	Studies in Veterinary Medicine, University of Berne
Aug 2001	Matura with main focus on Latin and English, Literargymnasium Kirchenfeld, Berne

Working Experience

Apr 2008 – present	Assistant position at the University of Zurich: • Teaching the organic chemistry practical course for undergraduate students • Supervision of lab technician apprentices • Mentoring of a master student during her master thesis
Jun 2006 – Sep 2006	Summer Student, Laboratory of Radio- and Environmental Chemistry, Paul Scherrer Institute, Villigen, CH: • Ion chromatographical measurements of an icecore from the Siberian Altai • Data analysis for climate reconstruction

Other activities

Dec 2009 – Jun 2011	Member of the organisation committee for the annual PhD-student day, Organic Chemistry Institute, University of Zurich
May 2005 – Dec 2007	Student representative in the department's leader's meetings, University of Bern
Dec 2003 – Dec 2007	Leader resort finances of the chemistry and biochemistry students' representative body (Fachverein Chemie und Biochemie), University of Berne

Achievements and Honors

Jun 2012	Swiss National Science Foundation (SNF) Post-doctoral Fellowship Grant for Postdoctoral Research at Stanford University (Prof. Dr. E. Kool), Palo Alto, USA
Jan 2011	Poster price for best group poster presentation, retreat of the CMSZH, Randa, CH
Jun 2009	Grant of the Nucleic Acid Center, University of Southern Denmark, for the participation in the Nucleic Acid Chemical Biology Summer School, Odense, Denmark

Presentations and Posters

Oral Presentation	Retreat of the CMSZH, Zuoz, February 2012
Posters	Fall Meetings of the Swiss Chemical Society (September 2008, September 2009, September 2010, September 2011)
	NACB Summer School, Odense, Denmark, June 2009
	EuCheMS Chemistry Conference, Nürnberg, September 2010
	Gordon Research Conference , Holderness, USA, June 2011

Publications

"Synthesis and Thermodynamic Studies of an Artificial *C*-Nucleotide as Precursor of Today's DNA", C. Roost, B. Bischof, J. S. Siegel, *manuscript in preparation*.

"Are Outer Ice Core Parts Usable for Analyses of Trace Species?", A. Eichler, C. Roost, T. Papina, M. Schwikowski, *Annual Report of the Laboratory for Radiochemistry and Environmental Chemistry of the Paul Scherrer Institute*, **2006**.